

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.6437 Seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453a-49

Perfect score: 1321

Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTLTGLKLLYEKNTTEKKG 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	975	73.8	254	2 F83660	hypothetical prote
2	900	68.1	233	2 S66100	conserved hypotet
3	855.5	64.8	259	2 AF1102	conserved hypotet
4	848.5	64.2	259	2 AF1464	conserved hypotet
5	755	57.2	273	2 E97293	probable transcrip
6	663	50.2	265	2 T36391	hypothetical prote
7	575	43.5	261	2 B87489	transcription acti
8	504.5	38.2	274	2 H86937	conserved hypotet
9	479.5	36.3	272	2 A70955	hypothetical prote
10	415.5	31.5	262	2 E75516	conserved hypotet
11	375.5	28.4	246	2 D72320	conserved hypotet
12	342.5	25.9	273	2 D71326	conserved hypotet
13	256.5	19.4	262	2 F70165	hypothetical prote
14	150	11.4	257	2 S75559	hypothetical prote
15	148.5	11.2	229	2 E70465	hypothetical prote
16	148	11.2	276	2 A12292	hypothetical prote
17	146	11.1	592	2 B81009	BlrA protein/Bvg a
18	146	11.1	592	2 H82031	probable biotin-fa
19	140.5	10.6	267	2 I40327	baf protein - Bord
20	123.5	9.3	248	2 H83111	hypothetical prote
21	115	8.7	242	2 A82637	conserved hypotet
22	105.5	8.0	597	2 B69251	probable electron
23	105	7.9	224	2 A99571	conserved hypotet
24	103.5	7.8	512	2 A98352	hydantoinase homol
25	103.5	7.8	512	2 AE2930	hydantoinase A [im
26	99.5	7.5	223	2 G71887	hypothetical prote
27	97	7.3	434	2 E96961	membrane protein c
28	96	7.3	394	2 AB1068	probable membrane
29	95	7.2	5825	2 T12117	polyprotein - fava

30 94 7.1 153 2 A95928  
31 94 7.1 467 2 AE3142  
32 94 7.1 520 2 G98145  
33 93.5 7.1 443 2 E88343  
34 93.5 7.1 465 2 T26686  
35 93 7.0 422 2 AD0484  
36 88.5 6.7 818 2 T15803  
37 88 6.7 425 2 T43861  
38 87.5 6.6 223 2 F64627  
39 87 6.6 473 2 C64371  
40 86.5 6.5 578 2 T03475  
41 86 6.5 267 2 D70819  
42 85 6.4 323 2 A70324  
43 85 6.4 503 2 D70930  
44 84.5 6.4 366 2 AH0939  
45 84.5 6.4 758 2 T39210

#### ALIGNMENTS

##### RESULT 1

F83660

hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: F83660

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: F83660

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <S>C>

A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0086

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 73.8%; Score 975; DB 2; Length 254;  
Best Local Similarity 70.8%; Pred. No. 9.6e-78;  
Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MIFVLDVGNNTVLGVYDGDDELKHHWRLETSRSKTEDEYGMKALLNHHVGLQFSIRGI 60  
Db 1 MILVIDGNTNTVLGVYDDETLVHHRLATSRQKTEDEYAMTVRSFLDFHAGLQFQIDGI 60  
Qy 61 IISVWPPIMPALERMLKYPHFKPLIVGPGIKTGLDIKYNPNRVEGADRIYNAVAGIHL 120  
Db 61 VISSVPPMPSLEOMCKKYPHVPMLIGPGIKTGLNLYKYNPKVEGADRIYNAVAGIEL 120  
Qy 121 YGSPLIIVDFGTATTCYINHKQYMGGAIAPIGMISTEALFARAALPRIETARPDDII 180  
Db 121 YGPAIVVDFGTATTCYINHKQYMGGAIAPIGMISTEALYHRASKLPRIETAKPKQV 180  
Qy 181 GKNTVSAMQAGILYGVQVEGIYVRMKAKSKIPKPKVIATGGLAPLIIASESDIIDVDFP 240  
Db 181 GTNTIDSMQSGIFYGVSDVGVVYKRMKAQAESEPKVIATGGLAKLIGTESETIDVDSF 240  
Qy 241 LTLTGLKLLYEKN 253  
Db 241 LTLKGLQLIYEKN 253

##### RESULT 2

S66100

conserved hypothetical protein yacB - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000

C:Accession: S66100; E69740

R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994  
A:Title: Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome  
A:Reference number: S65967; MUID:96051385; PMID:7584024  
A:Accession: S66100  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <OGA>  
A:Cross-references: EMBL:D26185; NID:9467326; PIDN:BAA05305.1; PID:dl005847; PID:9467459  
A:Note: the nucleotide sequence was submitted to the EMBL data library, December 1993  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Broutillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, P.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, Y.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
-keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69740  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <KUN>  
A:Cross-references: GB:299104; GB:AL009126; NID:92632267; PIDN:CAB11846.1; PID:el182003;  
A:Experimental source: strain 168  
A:Genetics: yacB  
A:Gene: yacB  
A:Start codon: TTG  
A:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 68.1%; Score 900; DB 2; Length 233;  
Best Local Similarity 78.2%; Pred. No. 3.1e-71;  
Matches 169; Conservative 27; Mismatches 20; Indels 0; Gaps 0;  
QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
DB 1 MLVLDVGNNTVLGVYDGDGKLEYHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
DB 1 MLVLDVGNNTVLGVYDGDGKLEYHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
DB 1 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
QY 121 YGSPLIIVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALFARAALPRIETARPDDII 180  
DB 121 YGNPLIVVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALYRAAKLPRIETARPDDII 180  
QY 181 GKNVTSAMOAGILYGVGVQVEGIVSRMKAKSKIPP 216  
DB 181 GKNVTSAMOAGILYGVGVQVEGIVSRMKAKSKIPP 216  
RESULT 3  
AF1102  
conserved hypothetical protein lmo0221 [imported] - *Listeria monocytogenes* (strain EGD-e  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
R:Accession: AF1102  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1102  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl6409586; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0221  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 64.8%; Score 855.5; DB 2; Length 259;  
Best Local Similarity 62.9%; Pred. No. 2.8e-67;  
Matches 163; Conservative 41; Mismatches 52; Indels 3; Gaps 1;  
QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
DB 1 MLVLDVGNNTVLGVYDGDGKLEYHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
DB 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
QY 121 YGSPLIIVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALFARAALPRIETARPDDII 180  
DB 121 YGTPVIVVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALYRAAKLPRIETARPDDII 180  
QY 181 GKNVTSAMOAGILYGVGVQVEGIVSRMKAKSKIPP 216  
DB 181 GKNVTSAMOAGILYGVGVQVEGIVSRMKAKSKIPP 216  
QY 241 LTLTKLILYKKN---TEK 256  
DB 241 LTLTKLILYKKN---TEK 256  
RESULT 4  
AF1464  
conserved hypothetical protein lin0253 [imported] - *Listeria innocua* (strain Clip124  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
R:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloer  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weh  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 64.2%; Score 848.5; DB 2; Length 259;  
Best Local Similarity 62.2%; Pred. No. 1.1e-66;  
Matches 161; Conservative 42; Mismatches 53; Indels 3; Gaps 1;  
QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
DB 1 MLVLDVGNNTVLGVYDGDGKLEYHWRITSRKSTDEYGMKIKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
DB 61 IISVVPIMPALERMCLYKPHIKPLIVGPGIKGLDIKYNPREVGADRIYNVAGIHL 120  
QY 121 YGSPLIIVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALFARAALPRIETARPDDII 180  
DB 121 YGTPVIVVDFGATATTCYCYINEHKOYMGGAIAPIGIMISTEALYRAAKLPRIETARPDDII 180





## H86937

conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: H86937  
R: Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hoo  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: H86937  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <STO>  
A:Cross-references: GB:AL450380; NID:gl3092576; PIDN:CAC29740.1; GSPDB:GN00147  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.2%; Score 504.5; DB 2; Length 274;  
Best Local Similarity 38.7%; Pred. No. 1.4e-36;  
Matches 105; Conservative 58; Mismatches 85; Indels 23; Gaps 6;  
QY 1 MIFVLDVGNNTVLGVGDGDELKHH-----WRIETSRKTEDEYGMKALLNHVGLQF 54  
Db 1 MLLAIDVRNTHVTVGLLSG--KREHAKVVOQWRIRTESEVTADELALIDGL--IGDGS 55  
QY 55 SDIRGIITSSVVPPIFMFALERMLCKYFIHKP-LIVGPGIKTGLDIDKYNPNREVGADRIVN 113  
Db 56 ERLAAGALSIVPSVLHEVRIMLDQYWPSPVPHVIEPGVGTGIPLLVDNPNKEVGADRIVN 115  
QY 114 AVAGIHLYGSLPIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAAPLRIE 173  
Db 116 CLAAHFHFGQAAIVDFGSSICVDWSAKGEFLGGAIAFGVQVSSDAAAARSALRRVEL 175  
QY 174 ARPDIDIGKNTVSAQAGILYGVQVEGIVSRMKA-----SKIPKVIATGGLAPLIA 224  
Db 176 ARPSRVVGNKNTVECMQAGVFGAGLVGDLGRMRQDVEEFSGDLGNRV--AVVATGHTA 233  
QY 225 PLIASDSIDVDPPELTLTGLKLLYEKNT 255  
Db 234 PLLLPETHVDHYDRHLLTGLRLVFERNLE 264

## RESULT 9

A70955  
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70955  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70955  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-272 <COL>  
A:Cross-references: GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv3600c  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 36.3%; Score 479.5; DB 2; Length 272;  
Best Local Similarity 37.8%; Pred. No. 2.2e-34;  
Matches 101; Conservative 56; Mismatches 93; Indels 17; Gaps 5;

QY 1 MIFVLDVGNNTVLGVGDGDELKHH-----WRIETSRKTEDEYGMKALLNHVGLQF 54  
Db 1 MLLAIDVRNTHVTVGLLSG--KREHAKVVOQWRIRTESEVTADELALIDGL--IGDGS 55  
QY 55 SDIRGIITSSVVPPIFMFALERMLCKYFIHKP-LIVGPGIKTGLDIDKYNPNREVGADRIVN 113  
Db 56 ERLTCTAALSTVPSVLHEVRIMLDQYWPSPVPHVIEPGVGTGIPLLVDNPNKEVGADRIVN 115  
QY 114 AVAGIHLYGSLPIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAAPLRIE 173  
Db 116 CLAAHDFRKAIAIVDFGSSICVDVVSAGKEFLGGAIAFGVQVSSDAAAARSALRRVEL 175  
QY 174 ARPDIDIGKNTVSAQAGILYGVQVEGIVSRMKA-----KSKIPKVIATGGLAPLIA 228  
Db 176 ARPSRVVGNKNTVECMQAGVFGAGLVGDLGRIRREDVSGFSVDHDAIVATGHTAPLL 235  
QY 229 SESDIDVDPPELTLTGLKLLYEKNT 255  
Db 236 PELHVDHYDQHLTLGLRLVFERNLE 262

## RESULT 10

E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain RL)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WH>  
A:Cross-references: GB:AE001905; GB:AE005513; NID:g6458144; PIDN:AAF10040.1; PID:g64:  
A:Experimental source: strain RL  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 31.5%; Score 415.5; DB 2; Length 262;  
Best Local Similarity 35.2%; Pred. No. 8.2e-29;  
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;

QY 2 IFVLDVGNNTVLGVGDGDELKHH-----WRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60  
Db 6 LLAVDIGNTTVLGLADASGALTHTWIRTNREMLPDDIALQLHGLFTLAGAPIP--RAA 63  
QY 61 IISVVPI-----MFALERMLCKYFIHKP-LIVGPGIKTGLDIDKYNPNREVGADRIVNAVA 116  
Db 64 VLSSVAPPVGENYALALKR-----HPMIDAFSAENLPDVTVELDPGSGVADRILCN--- 116  
QY 117 GIHLVYGS-----LIIVDFGTATTCYINEHKQYMGGAIAFGIMISTEALFARAAPL 168  
Db 117 ---LEGAERYLGLLDYAVVDEGTSTNEDVVGRRFLGILATGAQVSDALFARAAPL 173  
QY 169 PRIETARDDIDIGKNTVSAQAGILYGVQVEGIVSRMKAISKIPKVIATGGLAPLIA 228  
Db 174 PRITLQAPETAIGKNTVHALOSGLVFGAEMVDGLLRIRIRABELPGEAVAVATGFGFRTVQ 233  
QY 229 SESDIDVDPPELTLTGLKLLYEKNT 255  
Db 234 GICQEDVYDETTLRGLVFWASRSE 260

## RESULT 11

D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: D72320  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: D72320  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-246 <ARN>  
A;Cross-references: GB:AE001754; GB:AE000512; NID:94981417; PIDN:AAD35964.1; PID:g498141  
A;Experimental source: strain MSB  
C;Genetics:  
A;Gene: TW0883  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.4%; Score 375.5; DB 2; Length 246;  
Best Local Similarity 37.5%; Pred. No. 2.4e-25;  
Matches 94; Conservative 48; Mismatches 92; Indels 17; Gaps 9;  
QY 1 MFVLVDGNTNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALNHNHVLQFSDIRGI 60  
D 1 MYLLVDGNTHSVFSTEDCKTFRWLSGVQTEDELFSHLHPLL--GDAMREIKGI 57  
QY 61 IISVVPPIFALERMCLKYPFHKPLIVGPGIKTGLDIKYD--NPREGADRIVNAVAGI 118  
D 58 GVASVVTQNTVIERFSQKYFHSPIW--KAKNGC-VKWNKPNPSEVGADRVANVAVF 114  
QY 119 HLYGSPLLIYDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRTEIAPDD 178  
D 115 KEYKNGIILDMGTATTDVLV-VNGSEYGAIIPLGFFMVSLSFRGAKPLPLEVVEKPAF 173  
QY 179 IIGKNTVSAMQAGILYGVQVEGIVSRMK-ARKSKIPPVIATGGLAPLASESDII--D 235  
D 174 VVGKTEENIRLGVNGSVYALSGIIGRIKVEYGDLP--VVLGGQSKIV---KDMIKHE 228  
QY 236 VDPDFLTLTGL 246  
D 229 IFDEDLTIKGV 239

RESULT 12  
D71326  
Conserved hypothetical protein TP0431 - syphilis spirochete  
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
;Accession: D71326  
R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDoc they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A;Reference number: A71250; MUID:98332770; PMID:9665876  
A;Accession: D71326  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-273 <COL>  
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271  
A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0431  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 25.9%; Score 342.5; DB 2; Length 273;  
Best Local Similarity 31.0%; Pred. No. 2.1e-22;  
Matches 78; Conservative 60; Mismatches 105; Indels 9; Gaps 4;  
QY 2 IFVLVDGNTNTVLGVYDGDDE----LKHWRIFETSRKTEDEYGMKALNHNHVLQFSD 56  
D 1 MLLIDVGNHSHVFGI--QGENGRVGVCFRLFAPDARKTQDEYSLIHALCERAGVGRAS 59

QY 57 IRGIISSVPPIMFALERMCLKYPFHKPLIVGPGIKTGLDIKYDNP--REVGDRIVNA 114  
D 50 LRDAFISSVPPVLTITADAVAQISGVQVPPVPGWAYEHLVPRIPVRAEIGTDLVANA 119  
QY 115 VAGIHLGSPLLIYDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRTEI 174  
D 120 VAAVHFSAKCVVVDGTFALTFTAVDTGTLGIVGVAIPGLRTAVOSLHTGTQALPLVPLA 179  
QY 175 RPDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPPVIATGGLAPLASESDII 234  
D 180 LPDSVLGKDTTHAVQAGVVGFLFIRAMIAOCQELGRCACAAVITGGLSRLESSEVD-F 238  
QY 235 DVDPDFLTLTGL 246  
D 239 PPIDAQLTLTSLG 250

RESULT 13  
F70165  
Conserved hypothetical protein BB0527 - Lyme disease spirochete  
C;Species: Borrelia burgdorferi (Lyme disease spirochete)  
C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C;Accession: F70165  
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wison, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997  
A;Authors: Smith, H.O.; Venter, J.C.  
A;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A;Reference number: A70100; MUID:98065943; PMID:9403685  
A;Accession: F70165  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-262 <KLE>  
A;Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g26  
A;Experimental source: strain B31  
Query Match 19.4%; Score 256.5; DB 2; Length 262;  
Best Local Similarity 26.3%; Pred. No. 6.6e-15;  
Matches 67; Conservative 59; Mismatches 108; Indels 21; Gaps 5;  
QY 4 VLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALNHNHVLQFSDIRGIIS 63  
D 10 IIDIGNTSIAPALFKDQVNLFIKMTNMLRYDEVYSFFEENFD-----NVNKFVIS 63  
QY 64 SVVPPIMFALERMCLKYPFHKPLIVGPGIKTGLDIKYD---NPREGADRIVNAV 115  
D 64 SVVPIILNETFKNVIFSEFKIKPLFI-----GFDLNYDLTFNPYKSDKFLGSDVFPANLV 117  
QY 116 AGIHLG-SPLIYDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRTEI 174  
D 118 AAIEYNSFENVLVVDLGTACTIFAVSRQDGLGGIINGPLINFSLLDNNAVLIKKFPI 177  
QY 175 RPDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPPVIATGGLAPLASESDII 234  
D 178 TPNNLLERTTSGVNSGLFYQYKYLIEGVYDIKQYKKFNLIITGGNADLILSLIEIE 237  
QY 235 DVDPDFLTLTGLKLL 249  
D 238 FIFNIHLTVGEVRIL 252

RESULT 14  
S75559  
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C;Accession: S75559  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yae DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S. A; Reference number: S74322; MUID:97061201; PMID:8905231  
A; Accession: S75559  
A; Status: nucleic acid sequence not shown; translation not shown  
A; Molecule type: DNA  
A; Residues: 1-257 <RAN>  
A; Cross-references: EMBL:D90911; GB:AB001339; NID:g1553083; PIDN:BAAL8120.1; PID:d101895  
A; Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C; Genetics:  
A; Start codon: GTG

RESULT 15  
E70465  
hypothetical protein aq\_1924 - Aquifex aeolicus  
C;Species: Aquifex aeolicus  
C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C;Accession: E70465  
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MID:98196666; PMID:9537320  
A;Accession: E70465  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1229 <AQF>  
A;Cross-references: GB:AE000763; NID:g2984178; PID:AAC07720.1; PID:g2984188; GB:AE000655  
A;Experimental source: strain VF5  
C;Genetics:  
A;Gene: aq\_1924

```

- Query Match      11.2%  Score 148.5;  DB 2;  Length 229;
  Best Local Similarity 23.9%;  Pred. No. 1.5e-05;
  Matches          60;  Conservative 43;  Mismatches 111;  Indels 37;  Gaps 9;

QY      5 LDVGNNTVGLGVDDGDELKHHWRIETSRSKTEDEYGMKIALINHVLQFSD-IRGLIIS 63
      :||||: : :||: :||: :||: :||: :||: :||: :||:
Db      6 VDVGNSVDIALWEGKGVKDF--LKL SHEEFLKEEFPKALGATSVKQSEKVRG--- 59

QY      64 SVVPPIMFALERCLXYFHKPLIVPGIKTGLDIDKVDNPREVGCADRVINAVAGIHLGYS 123
      :||: :||: :||: :||: :||: :||: :||: :||:
Db      60 -KIPKIKFLKKE-----NFPIQVDYKTPETLGTDRVALAYSAKKFYKG 101

QY      124 PLIIVDFGATTTCYINEHKQYVGGAIAPGIMISTEALFAAAKLPRIEIARPDIDIIGN 183
      :||: :||: :||: :||: :||: :||: :||: :||:
Db      102 NVVVISAGTALVTDVLVEGK-FKGGFTLGLGKKLKILSLDLAGIEPPFEVEIFLGRS 160

QY      184 TVSAMQAGILYGVGVQGVIGIVSRMAKAKSI---PPKVIATGGIAPLATESDIDVVDVDF 240
      :||: :||: :||: :||: :||: :||: :||: :||:
Db      161 TRECVLGG--AYRESTEFIKSTLKLWRKVFKKFKVITGGEGKIFYSK---FGIYDPL 213

QY      241 LTLTGLK-LLY 250
      :||: :||: :||: :||: :||: :||: :||: :||:
Db      214 LVHGRGMRNLLY 224

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.18991 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-49

Perfect score: 1321

Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTLGLKLLYEKNTEKKG 258

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892.seqs, 41476328 residues

total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	900	68.1	233	1 YACB_BACSU	P37564 bacillus su
2	140.5	10.6	267	1 BAF_BORPE	Q45338 bordetella
3	87	6.6	425	1 GSA_CLOPE	Q32808 clostridium
4	87	6.6	473	1 AK_METJA	Q57991 methanococc
5	86	6.5	267	1 HIS6_MYCTU	Q53908 mycobacteri
6	86	6.5	303	1 V212_FOWPV	Q91523 fowlpox vir
7	85	6.4	304	1 HEM3_AQUAE	O66621 aquifex aeo
8	84.5	6.4	758	1 LEU2_SCHPO	O42899 schizosacch
9	84	6.4	361	1 OP22_HAEIN	P20149 haemophilus
10	84	6.4	419	1 Y092_HAEIN	Q57493 haemophilus
11	83.5	6.3	469	1 LEU2_RHILO	Q98631 rhizobium l
12	83	6.3	365	1 HHP1_SCHPO	P40235 schizosacch
13	83	6.3	401	1 CHRA_ALCEU	P17551 alcaligenes
14	83	6.3	421	1 CIMA_METJA	Q57878 methanococc
15	82.5	6.2	420	1 YD43_METJA	Q58739 methanococc
16	82	6.2	336	1 GLPX_ECOLI	P28860 escherichia
17	81.5	6.2	258	1 HIS6_CORGL	O31139 corynebacte
18	81.5	6.2	429	1 AROA_METJA	Q57925 methanococc
19	81.5	6.2	469	1 LEU2_AGRF5	Q86979 agrobacteri
20	81.5	6.2	494	1 HK25_YEAST	P29295 saccharomyc
21	81	6.1	424	1 SAHH_METKA	P58855 methanopyru
22	81	6.1	543	1 CH60_CLOPE	P30717 clostridium
23	81	6.1	610	1 FTSL_HAEIN	P45059 haemophilus
24	81	6.1	1184	1 POL2_GFLY	P48474 grapevine f
25	80.5	6.1	470	1 LEU2_AZOV1	P96195 azotobacter
26	80.5	6.1	579	1 YD49_MYCTU	Q11019 mycobacteri
27	80	6.1	359	1 OP21_HAEIN	P43839 haemophilus
28	80	6.1	788	1 PUR2_RHLE	Q99148 y bifunctio
29	79.5	6.0	469	1 LEU2_RHLE	Q92176 rhizobium m
30	79	6.0	1042	1 CARB_HALN1	Q9hp43 halobacteri
31	79	6.0	4451	1 GRSB_BACBR	P14688 b gramicidi
32	78.5	5.9	232	1 UPS_PYRAB	Q9v0k1 pyrococci
33	78.5	5.9	410	1 RAA2_CHLRE	Q9smh4 chlamydomon

34	73.5	5.9	478	1 GLN1_MYCTU	Q10377 mycobacteri
35	73.5	5.9	485	1 SAHH_STRCO	Q9kxm1 streptomyc
36	73.5	5.9	543	1 CH60_BARBA	P35635 bartonella
37	73.5	5.9	859	1 YD48_MYCTU	Q11018 mycobacteri
38	78	5.9	296	1 HIS1_YARLI	Q99145 yarrowia l1
39	78	5.9	360	1 OP26_HAEIN	Q48216 haemophilus
40	78	5.9	585	1 CH60_PYRSA	P46224 pyrenomonas
41	78	5.9	1146	1 MMLC_MYCTU	Q50585 mycobacteri
42	77.5	5.9	261	1 HIS6_MYCLE	Q9x7c2 mycobacteri
43	77.5	5.9	439	1 MAO1_BACSU	P54572 bacillus su
44	77.5	5.9	469	1 LEU2_BRUME	Q8yjc9 brucella me
45	77.5	5.9	502	1 HAK2_ARATH	P93834 arabidopsis

#### ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowaka A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
"The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC the European Bioinformatics Institute and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D26185; BAA05305.1; -  
 DR EMBL; Z99104; CAB11846.1; -  
 DR Subtilist; BG10133; yacB.  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs; TIGR00671; baf; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match 68.1%; Score 900; DB 1; Length 233;  
 Best Local Similarity 78.2%; Pred. No. 5.2e-70;  
 Matches 169; Conservative 27; Mismatches 20; Indels 0; Gaps 0;

1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60  
 1 MLLVLDVGNNTVLGVYDGDDELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60  
 61 IISVVPPIPMFALERMCKLYFHFKPLIVGPGIKTGLDKNPREVGADRVNAVAGIHL 120  
 61 IISVVPPIPMFALERMCKLYFHFKPLIVGPGIKTGLDKNPREVGADRVNAVAGIHL 120  
 121 YGSLIIVDGTATTCYINEHKOYMGATAPGIMISTEALFAAKLPRIEIRAPDII 180  
 121 YGSLIIVDGTATTCYINEHKOYMGATAPGIMISTEALFAAKLPRIEIRAPDII 180  
 181 GKNTVSAMQAGILYGVQGEVIGYVRMKAKSKIPP 216  
 181 GKNTVSAMQAGILYGVQGEVIGYVRMKAKSKIPP 216

RESULT 2  
 BAF\_BORPE STANDARD; PRT; 267 AA.  
 AC Q45338; Q45373;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Bvg accessory factor.  
 GN BAF.  
 QS Bordetella pertussis.  
 Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 Bordetella.  
 NCBI\_TaxID=520;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BP504;  
 RX MEDLINE-953525323; PubMed-7601846;  
 RA Deshazer D., Wood G.E., Friedman R.L.;  
 RT "Identification of a Bordetella pertussis regulatory factor required  
 RT for transcription of the pertussis toxin operon in Escherichia  
 RT coli."  
 RL J. Bacteriol. 177:3801-3807(1995).  
 RN [2]  
 SEQUENCE OF 1-38 FROM N.A.  
 RC STRAIN-BP504;  
 RA Wood G.E., Friedman R.L.;  
 RT "Identification of a bira homolog in Bordetella pertussis."  
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE OF 239-267 FROM N.A.  
 RP STRAIN-BP536;  
 RC MEDLINE-96419162; PubMed-8821935;  
 RA Allen A.G., Maskell D.J.;  
 RT "The identification, cloning and mutagenesis of a genetic locus

RT required for lipopolysaccharide biosynthesis in Bordetella  
 RT pertussis."  
 RL Mol. Microbiol. 19:37-52(1996).  
 CC FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
 CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
 CC RNA POLYMERASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U12020; AAA75361.1; -  
 DR EMBL; AF016461; AAC68834.1; -  
 DR EMBL; X90711; CA62242.1; -  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 KW Transcription regulation; Activator.  
 SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 10.6%; Score 140.5; DB 1; Length 267;  
 Best Local Similarity 24.7%; Pred. No. 6.1e-05;  
 Matches 66; Conservative 42; Mismatches 106; Indels 53; Gaps 12;

1 MIFVLDVGNNTVLGVYDGDDELKHHWRIETSRKTE-----DEYGMKIKAL----- 46  
 1 MLLIDSGNSRLKVGWDFDP-----APQAREPAPAFDNLDDALGRWLATLRPROR 54  
 47 ---LNHVGLQFSDIRGIISVVPPIPMFALERMCKLYFHFKPLIVGPGIKTGLDKNP 103  
 55 ALGVNVAGL-----ARGEIAATL-----RAGGCDIRWLRAQPLAM--GLRNG---YRNP 99  
 104 REVGADR---IVNAVAGIHLVGSPLIIVDFGTATTCYINEHKOYMGATAPGIMISTEA 160  
 100 DQLGADRWACWGVGLARQPSVHPPLLVASFGTATTLDITGPDNVPFGGLPGAMMGA 159  
 161 LFARAAKLPRIEIARPPDDIIGKNTVSAMQAGILYGVQGV---EGIVSRMKAKSKIPP 218  
 160 LAYGTAHLPLADGLVADYPI--DTHQATASGIAAAGAIIVQWLGRQRYGO--APEIV 215  
 219 ATGGLAPLIASESDIIVDVFLLTGT 245  
 216 VAGGWPVEVQEAERL-----LAVTG 236

RESULT 3  
 GSA\_CLOPE STANDARD; PRT; 425 AA.  
 AC Q92NC8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)  
 GN HEML OR CPE1432.  
 OS Clostridium perfringens.  
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 NCBI\_TaxID=1502;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN-NCTC 8237;  
 RA MEDLINE-20049836; PubMed-10585141;  
 RA Koyama M., Katayama S., Kaji M., Taniguchi Y., Matsushita O.,  
 RA Minami J., Morita S., Okabe A.;  
 RT "A Clostridium perfringens hem gene cluster contains a cysG(B)  
 RT homologue that is involved in cobalamin biosynthesis."  
 RL Microbiol. Immunol. 43:947-957(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN=13 / Type A;
RX PubMed-11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -|- CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate = 5-
CC aminoolevulinic acid.
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -|- PATHWAY: Porphyrin biosynthesis by the C5 pathway; second step.
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -|- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EXBL: AB017186; BAA74784.1;
CC EMBL: AP003190; BAB81138.1; ALT_INIT.
CC HSP: P24630; 2GSA.
CC InterPro: IPR000954; Aminotran_3.
CC Pfam: PF00202; aminotran_3; HemL.
CC TIGRFAMs: TIGR00713; hemL; 1.
CC PROSITE: PS00600; AA_TRANSFER_CLASS_3; 1.
CC Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate;
CC Complete proteome.
CC BINDING 265 265 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC CONFLICT 73 73 E -> K (IN REF. 1).
CC CONFLICT 323 323 N -> D (IN REF. 1).
CC CONFLICT 356 356 I -> M (IN REF. 1).
CC CONFLICT 422 422 T -> S (IN REF. 1).
CC PROSITE: TIGR00713; hemL; 1.
CC PROSITE: PS00600; AA_TRANSFER_CLASS_3; 1.
CC PROSITE: B23DFD62CE39B9E2 CRC64;
CC SEQUENCE 425 AA; 46881 MW; B23DFD62CE39B9E2 CRC64;
KW Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate;
KW Complete proteome.
FT BINDING 265 265 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT CONFLICT 73 73 E -> K (IN REF. 1).
FT CONFLICT 323 323 N -> D (IN REF. 1).
FT CONFLICT 356 356 I -> M (IN REF. 1).
FT CONFLICT 422 422 T -> S (IN REF. 1).
SQ SEQUENCE 425 AA; 46881 MW; B23DFD62CE39B9E2 CRC64;

Query Match 6.6%; Score 87; DB 1; Length 425;
Best Local Similarity 21.8%; Pred. No. 4;
Matches 42; Conservative 32; Mismatches 41; Indels 78; Gaps 12;

QY 11 NTVLGVYDGDDELKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGIISVW---- 66
b 175 NTLGIYNDER-----QVE--ELFEKYG-----NDIAGIIIEPVAGNMG 211

QY 67 -----PRIMFALEKRLKY-----FHKP-----LIVGPGI 92
212 VVKCDPRFMKRLKELCDKYGALLIFDEVMCGFRVAYKGAQTLFDVKPLVYAKIMGGGL 271

QY 93 KTGLEDIYDNPREGADRIYNAVAGIH----LYGSPLIIVDFGTATTY-----CYINEH 142
Db 272 PCG---AYGGRREIMEN--LSPLGGVYQAGTMSGNP-IVMSAGLATVKKLYENPSYIN-H 324

QY 143 KQYMGAIAPGIM 155
Db 325 IEKIGSKLEKGV 337

RESULT 4
AK_METJA STANDARD; PRT; 473 AA.
AC Q57991;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable aspartokinase (EC 2.7.2.4) (Aspartate kinase).
GN M30571.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

```

```

OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -|- CATALYTIC ACTIVITY: ATP + L-aspartate = ADP + 4-phospho-L-
CC aspartate.
CC -|- PATHWAY: FIRST STEP IN THE COMMON BIOSYNTHETIC PATHWAY LEADING
CC FROM ASP TO DIAMINOPIMELATE AND LYS, TO MET, TO ILE AND TO THR.
CC -|- SIMILARITY: BELONGS TO THE ASPARTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67506; AAB98565.1;
CC TIGR: M30571;
CC InterPro: IPR002912; ACT.
CC InterPro: IPR001048; Aa_kinase.
CC InterPro: IPR005260; Asp_kin_monofn.
CC InterPro: IPR001341; Aspartate_kinase.
CC Pfam: PF00696; aakinas; 1.
CC Pfam: PF01842; ACT; 2.
CC TIGRFAMs: TIGR00656; asp_kin_monofn; 1.
CC TIGRFAMs: TIGR00657; asp_kinases; 1.
CC PROSITE: PS00324; ASPARTOKINASE; 1.
CC Transferrase; Kinase; Threonine biosynthesis; Complete proteome.
KW SEQUENCE 473 AA; 51392 MW; 811C6E0F4B66BC5F CRC64;

Query Match 6.6%; Score 87; DB 1; Length 473;
Best Local Similarity 23.3%; Pred. No. 4.5;
Matches 67; Conservative 43; Mismatches 98; Indels 80; Gaps 14;

QY 35 TEDEY-----GMKALLNHVGLQ-----FSDIRGIISV--VVP-----PIMPALE 74
Db 197 TEEGYITTLGRGSDYSAAALIGVGLDADIIEIWTVDVSGVYTPRLVPTARRIPKLSYIE 256

QY 75 RCLKYFHFK---PLIVGPGIKYGLDIKYDN---PREVG-----ADRIYNA----- 114
Db 257 AMELAYFGAKVLRHTIEPAMEKGPILVKNTEPESEGTLLITNDMEMSDSIYKAISTIK 316

QY 115 -VAGIHLGSPLIIVDFGTATTYCYINEHKQYV-----GGAIPAGTMISTEALFARAAL 168
Db 317 NVALINIFGAGWGVSGTAAARIFKALGEEVNVILISQSSNETNISLVSEEDVDKAKA 376

QY 169 PRIETARPDDIIGKNT-----VSAMQAGI--LYGYGVQGVGVISVRMK 208
Db 377 LKREFGD---FGKKSFLNNLIRDVSVKDVCKVISVVGAGMRGAKGIAGKIFTAVSESG 432

QY 209 AKSKIPPVKVIATGGLAPLIASESDIIDVVDPLTLTGLKLLYEKNTKEK 256
Db 433 ANI-----KMAIQG-----SSEVNISFVIDERDLNLCVRKLHEKFTIEK 470

RESULT 5
HIS6_MYCTU
ID HIS6_MYCTU STANDARD; PRT; 267 AA.

```





РЕСПИ.Т. 7

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/or\\_send\\_an\\_email\\_to\\_license@sib-sib.ch](http://www.isb-sib.ch/announcement/or_send_an_email_to_license@sib-sib.ch)).  
-----

Query Match 6.4%; Score 85; DB 1; Length 304;  
Best Local Similarity 21.0%; Pred. No. 4;  
Matches 49; Conservative 36; Mismatches 84; Indels 64; Gaps 8;

8 JUL 1958

RC STRAIN=972;

CC	-I- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC	and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC	-I- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
CC	H <sub>2</sub> O.

-I- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-isopropylmalate.  
CC  
CC  
CC  
-I- PATHWAY: Leucine biosynthesis; second step.  
CC  
CC  
-I- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
CC  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z99262; CAB16402.1; -  
 DR InterPro: IPR000573; Aconitase\_C.  
 DR InterPro: IPR001030; Aconitase\_N.  
 DR InterPro: IPR004430; Leuc.  
 DR InterPro: IPR004431; Leuc.  
 DR Pfam: PF00330; aconitase; 1.  
 DR Pfam: PF00694; Aconitase\_C; 1.  
 DR PRINTS; PR00415; ACONITASE.  
 DR ProDom; PD000511; Aconitase\_N; 1.  
 DR TIGRFAMs; TIGR00170; leuc; 1.  
 DR TIGRFAMs; TIGR00171; leuc; 1.  
 DR PROSITE; PS00450; ACONITASE.1; 1.  
 DR PROSITE; PS01244; ACONITASE.2; 1.  
 DR Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S.  
 DR METAL; 359 359 IRON-SULFUR (4Fe-4S) (BY SIMILARITY).  
 DR METAL; 420 420 IRON-SULFUR (4Fe-4S) (BY SIMILARITY).  
 DR METAL; 423 423 IRON-SULFUR (4Fe-4S) (BY SIMILARITY).  
 SQ SEQUENCE 758 AA; 82782 MW; CE78C36828380E47 CRC64;

Query Match 6.4%; Score 84.5; DB 1; Length 758;

Best Local Similarity 24.0%; Pred. No. 13;  
 Matches 37; Conservative 30; Mismatches 58; Indels 29; Gaps 7;

QY 102 NPREVADRIYNAVAGIHLVSPILIVDFGATTTCYINEHKQYMGGAIAFGI----- 154  
 DB 113 NDRQGVHVGIPGPGTLPCTTLVCGDSHTST-----HGAF--GALAFGIGTSEVH 163  
 QY 155 MISTEALFAAKLPRIEI--ARDDIIGKNTV-----SAMQAGILYGVGO-VBGI 203  
 DB 164 VLATQTILQRKSKNMIRVNGKLEPIASDKLLHLIGVIGTAGTGVSIFECGEAIGL 223  
 QY 204 VSRMKASKTPPKVIATGGLAPLASESDIIDV 237  
 DB 224 --SMEARSMCMNSIEAGARAGMIAPDTPEY 255

# RESULT 9

OP22\_HAEIN STANDARD; PRT; 361 AA.  
 AC P20149;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 Outer membrane protein P2 precursor (OMP P2).  
 OMP2.  
 Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Serotype B;  
 RX MEDLINE-89079316; PubMed-2535836;  
 RA Munson R.S. Jr., Tolan R.W. Jr.;  
 RT "Molecular cloning, expression, and primary sequence of outer  
 membrane protein P2 of Haemophilus influenzae type b.";  
 RL Infect. Immun. 57:88-94(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Serotype B;  
 RX MEDLINE-90158127; PubMed-2576096;  
 RA Munson R.S. Jr., Bailey C., Grass S.;  
 RT "Diversity of the outer membrane protein P2 gene from major clones of  
 Haemophilus influenzae type b.";  
 RL Mol. Microbiol. 3:1797-1803(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN-Serotype B;  
 RX MEDLINE-89173305; PubMed-2538396;  
 RA Hansen E.J., Hasemann C., Clausell A., Capra J.D., Orth K.,  
 RA Moonaw C.R., Slaughtor C.A., Latimer J.L., Miller E.E.;  
 RT "Primary structure of the porin protein of Haemophilus influenzae  
 type b determined by nucleotide sequence analysis.";  
 RL Infect. Immun. 57:1100-1107(1989).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC -!- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; J03359; AAA24993.1; -  
 DR EMBL; A09003; CAA00819.1; -  
 DR PIR; A30542; A30542.  
 DR PIR; S09622; S09622.  
 DR InterPro: IPR003229; OMP\_2.  
 DR InterPro: IPR001702; Porin-gram-ve.  
 DR Pfam; PF00267; Gram-ve\_porins; 1.  
 DR ProDom; PD000808; OMP\_2; 1.  
 DR Outer membrane; Transmembrane; Porin; Signal.  
 KW SIGNAL 1 20  
 FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.  
 SQ SEQUENCE 361 AA; 39701 MW; AF104C9CD8942D69 CRC64;

Query Match 6.4%; Score 84; DB 1; Length 361;  
 Best Local Similarity 23.0%; Pred. No. 6;  
 Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSRSKT-----EDEYGMW-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
 DB 119 VGLRAKTIADGITSADKEDYGVLLNNSDIPTSGNTVGYTFKGIDLVLGA---NYLLAQ 175  
 QY 74 ERMCLKYPHKLIVGVPGIKTGLIDIKYDNPREVGVADRIYNAVAGIHLVSPILIVDFGTA 133  
 DB 176 KREGAKGKPKNDKAGEVRIG--EINNGIQVGAKYDANDIVAKTAYG-----R 222  
 QY 134 TTYCY--INHKQYMGGAIAP-----GIMISTEALFAR 164  
 DB 223 TNYKNESDEHKQQLNGVLATLGYRFSGLGLVSLDSGYAK 263

# RESULT 10

Y092\_HAEIN STANDARD; PRT; 419 AA.  
 AC Q57493;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein HI0092.  
 GN HI0092.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Furumann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RL Rd";  
CC Science 269:496-512(1995).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U32694; AAC21770.1; -  
DR TIGR: HI0092;  
DR InterPro: IPR003474; GntP\_permease.  
DR Pfam: PF02447; GntP\_permease; 1.  
DR Hypothetical protein; Transmembrane; Complete proteome.  
W T TRANSMEM 1 21 POTENTIAL.  
T TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 56 86 POTENTIAL.  
FT TRANSMEM 101 121 POTENTIAL.  
FT TRANSMEM 174 194 POTENTIAL.  
FT TRANSMEM 216 236 POTENTIAL.  
FT TRANSMEM 242 262 POTENTIAL.  
FT TRANSMEM 280 300 POTENTIAL.  
FT TRANSMEM 311 331 POTENTIAL.  
FT TRANSMEM 349 369 POTENTIAL.  
FT TRANSMEM 396 416 POTENTIAL.  
SQ SEQUENCE 419 AA; 42781 MW; 6DCAS4BB0A55FF73 CRC64;

Query Match 6.4%; Score 84; DB 1; Length 419;  
Best Local Similarity 26.6%; Pred. No. 7.1;  
Matches 69; Conservative 34; Mismatches 94; Indels 62; Gaps 16;  
QY 29 ETSRSKTEDEYGMKAL- LNHVGFQDSIRGIISSVPPIMFALERC-LKYPH 83  
DB 87 ETYNKLEGTALLALATMLTAVGV-FVDVAVITVS- PALALSRSDLSKAAI 141  
QY 84 KPLVGPCKTGKLDIKYDNPREGADRVNAVAGIHLXGSPLIIVD- FGTAT 135  
DB 142 LLAMIGGG-KAG-NIMSPNP- NATAAADTFLPLTSVMAGIIPALGLIILT 190  
QY 136 YC- YINEHKQYMGGAIAFGIMISTEAL- FARAALKPRIEI- ARPDDITGKNTV 185  
DB 191 YFLAKRLINKSKVTDEK- IVLETONLPSFUTALVAPLVAIILLALRPLFDIKVDPL 247  
QY 186 SAMOAGILYG- YVGOVEGIVS-RMKAKSKIPP- KVIATGGLAPLIASESDIIDVDP 239  
DB 248 IALPLGLLIGAFCKGLNINSYAINGLSKWTPVAILMLLTGALLAGIITAN- 297  
QY 240 FLTLTGLKLYEKNTEKKG 258  
DB 298 - SGLKEVLIOGLEHSS 312

RESULT 11  
LEU2\_RHILO  
ID LEU2\_RHILO STANDARD; PRT; 469 AA.  
AC Q9BEF1;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 3-Isopropylmalate dehydratase large subunit (EC 4.2.1.33)  
DE (Isopropylmalate isomerase) (alpha-IPM isomerase) (IPMI).  
GN LEUC OR ML4272.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-WAF303099;  
RX MEDLINE-21082930; PubMed-11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
CC -!- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate  
CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.  
CC -!- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +  
CC H(2)O.  
CC -!- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-  
CC isopropylmalate.  
CC -!- PATHWAY: Leucine biosynthesis; second step.  
CC -!- SUBUNIT: Heterodimer of leucC and leucD (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 1  
CC SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AP003003; BAB50968.1; -  
DR InterPro: IPR001030; Aconitase\_N.  
DR InterPro: IPR004430; Leuc.  
DR Pfam: PF00330; aconitase; 1.  
DR PRINTS: PR00415; ACONITASE.  
DR ProDom: PD000511; Aconitase\_N; 1.  
DR TIGRfam: TIGR00170; leucC; 1.  
DR PROSITE: PS00450; ACONITASE\_1; 1.  
DR PROSITE: PS01244; ACONITASE\_2; 1.  
KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 350 350 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 469 AA; 50871 MW; 641532D3D4F06888 CRC64;  
Query Match 6.3%; Score 83.5; DB 1; Length 469;  
Best Local Similarity 24.8%; Pred. No. 8.9;  
Matches 38; Conservative 27; Mismatches 61; Indels 27; Gaps 6;  
QY 95 GLDIKYDNPREGADRVNAVAGIHLXGSPLIIVDFGTATYCYINEHKQYMGGAIAPCI 154  
DB 97 GVEYSENDIROGIVHIIGPEQGTLPGMTIVCGDSHTST- HGAF--GALAHI 147  
QY 155 -----MISTEALFARAAR--LPRIEARPDDIIGKNTVSAM-----QAGILYGV 197  
DB 148 GTSEVEHVLATQTLIORKAKNMLRVVDGQLPEGVTAADIIILAIIGEIGTAGTGVIIEYA 207  
QY 198 GQVEGIVSRMKAKSKIPPVATGGLAPIASE 230  
DB 208 GEAIRSLS-MEGRMTICNMSIEGAGARAGLIAAD 239  
RESULT 12  
HHP1\_SCHPO  
ID HHP1\_SCHPO STANDARD; PRT; 365 AA.  
AC P40235;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Casein Kinase I homolog hhp1 (EC 2.7.1.-).  
GN HHP1 OR SP3C3H7.15.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;

CC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SP66;  
RX MEDLINE=94354807; PubMed=8074660;  
RA Kearney P., Ebert M., Kuret J., and sequence analysis of two novel fission yeast  
RT "Molecular cloning and sequence analysis of two novel fission yeast  
RL casein kinase-1 isoforms.";  
RL Biochem. Biophys. Res. Commun. 203:231-236(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=94298768; PubMed=8026462;  
RX Dhillon N., Hoekstra M.F.;  
RA "Characterization of two protein kinases from Schizosaccharomyces  
RT pombe involved in the regulation of DNA repair.";  
RL EMBO J. 13:2777-2788(1994).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
RA Gencies S., Goble L., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds R., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs C., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gallard J., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RA "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -1- FUNCTION: INVOLVED IN DNA REPAIR. HAS A PROBABLE ROLE IN REPAIRING  
ALYLATED DNA AND MAY REGULATE THE ACTIVITY OF PROTEIN(S) INVOLVED  
IN DOUBLE STRAND BREAK REPAIR CAUSED BY GAMMA RAYS.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CASEIN KINASE I SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U10863; AAA21544.1; -  
CC EMBL; X78871; CAA55473.1; -  
CC EMBL; S4031261; CAA20311.1; -  
CC PIR; S44196; S44196.  
CC HSRP; 006486; ICKI.  
CC InterPro; IPR000719; Euk\_pkinase.  
CC InterPro; IPR002290; Ser\_thr\_pkinase.  
CC Pfam; PF00069; pkinase; 1.  
CC ProDom; PD000001; Euk\_pkinase; 1.  
CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding; DNA repair;  
KW Nuclear protein.  
FT DOMAIN 11 279 PROTEIN KINASE.  
FT NP\_BIND 17 25 ATP (BY SIMILARITY).  
FT BINDING 40 40 ATP (BY SIMILARITY).  
FT ACT\_SITE 130 130 BY SIMILARITY.  
FT DOMAIN 299 323 GLN/PRO-RICH.  
SQ SEQUENCE 365 AA; D637C95055F1339 CRC64;  
Query Match 6.3%; Score 83; DB 1; Length 365;  
Best Local Similarity 23.0%; Pred. No. 7.4; Gaps 11;  
Matches 51; Conservative 47; Mismatches 92; Indels 32;  
QY 16 VYDGDGLKHWRIETSRK---TEDEYGMKALLNHVGLQFSDIRGI-----IISVV 66  
DB 31 VVSGEVA--IKLESTRAKHPQLEYEY-RVYRILSGVGIPFVWFVFGVCDYNAMVMDLL 87  
QY 67 PPIFMALERMCLKYFIHKP-LIVPGIKTKGLDIDKYNP---REVGAADRIYNAVAGIHLYG 122  
DB 88 GPSLEDLFNCNKFSLTKVLLADQLISRIETHSKSFHLRDIKPD---NFLMGIGKRG 144  
QY 123 SPLIIVDFCTATYCYVINEHKOYMGGAIAPIGIMISTEALFAAKLPRIETARPDDLIIGK 182  
DB 145 NOVNIIDFGLAKY---ROHKTHLHPYRENKNTLTGTARYASINTHLIGESQRDDLES 201  
QY 183 NTVSAMQAGILXYGVGV---EGIVSRMKAK--SKIPPKVIAT 220  
DB 202 GIV-----LVVFCRSLPQWGLKATTKKQYKIMEKKIST 237  
RESULT 13  
CHRA\_ALCEU STANDARD; PRT; 401 AA.  
AC PL7551;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Chromate transport protein.  
GN CHRA.  
OS Alkaligenes eutrophus (Ralstonia eutropha).  
OG Plasmid pMOL28  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=510;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CH34;  
RX MEDLINE=90202806; PubMed=2180932;  
RA Nies A., Nies D.H., Silver S.;  
RT "Nucleotide sequence and expression of a plasmid-encoded chromate  
resistance determinant from Alkaligenes eutrophus.";  
RL J. Biol. Chem. 265:5648-5653(1990)  
CC -1- FUNCTION: THIS PROTEIN REDUCES CHROMATE ACCUMULATION AND IS  
ESSENTIAL FOR CHROMATE RESISTANCE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- INDUCTION: BY CHROMATE.  
CC -1- SIMILARITY: 20% IDENTITY TO PSEUDOMONAS AERUGINOSA CHROMATE  
TRANSPORT PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ313327; CAC42411.1; -  
CC PIR; B35177; B35177.  
CC InterPro; IPR003370; Chromate\_transp.  
CC Pfam; PF02417; Chromate\_transp; 2.  
CC TIGRfams; TIGR00937; 2A51; 1.

KW	Transport; Plasmid; Transmembrane; Chromate resistance.
FT	TRANSMEM 26 46 POTENTIAL.
FT	TRANSMEM 67 87 POTENTIAL.
FT	TRANSMEM 93 113 POTENTIAL.
FT	TRANSMEM 124 144 POTENTIAL.
FT	TRANSMEM 172 192 POTENTIAL.
FT	TRANSMEM 215 235 POTENTIAL.
FT	TRANSMEM 237 257 POTENTIAL.
FT	TRANSMEM 282 302 POTENTIAL.
FT	TRANSMEM 330 350 POTENTIAL.
FT	TRANSMEM 356 376 POTENTIAL.
FT	TRANSMEM 379 399 POTENTIAL.
SEQ	SEQUENCE 401 AA; 42642 MW; 6DC963FBADA6DBC5 CRC64;
Query Match 6.3%; Score 83; DB 1; Length 401;	
Best Local Similarity 23.8%; Pred. No. 8.2;	
Matches 46; Conservative 36; Mismatches 75; Indels 36; Gaps 10;	
QY	81 FHKPLVLVGIGIKTGDLIKVDNPREVGADRIIVNAGIHLYGSLIIVDF---GSTATYC 137
DB	207 FAATPLPAASGMMSTLD--WPLLSQIG---VFFAKAGAFVFGSLAIVPFLYGVVTEYH 261
QY	138 YINEKHQYMGGA---IAPG-IMISTEALFARAACLPRIBETARDDDIIGKNTVSAQA-- 190
DB	262 WLND-KQFDVAVAMTTPGVVITFGIYVAGLPGACVAAATFPCYLFVLPAPY 320
QY	191 ----GILYGVGVQGVGIVSRMAKSKIPPVKVIATGGL--APLIASESDIIDVDPFUTLT 244
DB	321 FKKYGLPAILAFVDGYTA-----AAGATGAVIVLAKRSIVDIPITALLIV 368
QY	245 GLKLL--YERKTE 255
DB	369 TVALLLKFKKLSE 381
RESULT 14	
CIMA_METJA	
ID	CIMA_METJA STANDARD; PRT; 491 AA.
AC	Q58787;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	(R)-citramalate synthase (EC 4.1.3.-).
CIMA	OR MJ1392.
GN	Methanococcus jannaschii.
OS	Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC	Methanocaldococcaceae; Methanocaldococcus.
NCBI_TaxID	=2190;
N	[1]
SEQUENCE	FROM N.A.
RP	STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RC	MEDLINE-96337999; PubMed-8688087;
RX	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glöck A.,
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA	Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus
RL	jannaschii.";
RL	Science 273:1058-1073(1996).
RP	[2]
SEQUENCE	OF 1-10, AND CHARACTERIZATION.
RX	PubMed-9864346;
RA	Howell D.M., Xu H., White R.H.;
RT	(R)-citramalate synthase in methanogenic archaea.";
RA	J. Bacteriol. 181:331-333(1999).
CC	-1- FUNCTION: Catalyzes the condensation of pyruvate and acetyl-
CC	coenzyme A to form (R)-citramalate.
CC	-1- SUBUNIT: Homodimer.
CC	-1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE

```
RL Science 273:1058-1073(1996).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -I- SIMILARITY: BELONGS TO THE AMT1/MEP/NRGA FAMILY OF AMMONIUM
CC TRANSPORTERS (TC 2.49).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67574; AAB999352.1; -
CC TIGR; MJ1343; -
DR InterPro; IPR001905; Ammonium transp.
DR Pfam; PF00909; Ammonium transp; 1.
DR TIGRFAMs; TIGR00836; amt; 1.
DR PROSITE; PS01219; AMMONIUM_TRANS; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 295 315 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
SQ SEQUENCE 420 AA; 43785 MW; F7AE2554D19DACA2 CRC64;

Query Match          6.2%; Score 82.5; DB 1; Length 420;
Best Local Similarity 23.8%; Pred. No. 9, 6;
Matches 36; Conservative 22; Mismatches 50; Indels 43; Gaps 7;

OY 145 YMGAIAPGI--MISTEALFARAALPRIEIARPDIIIGKNTVSAQAGIL-----YGY- 196
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 179 YAGSLVHVHGLGGLGALGAIAALGRIGRFVDCRPVPIILGHNPMAVFGAFALAIGWYGFN 238
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 197 -----VGQVSGIVSEMKAKSIPKPVATGGLAPLIASESDII----- 234
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 239 VGSSIALGDISGLVCATTMA-----MAGGGIGALIASRNDVLTANGIVAGLVAICSG 292
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 235 -DVVDPPFTL-----TGLKL-LYEKNTKKG 258
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 293 TDVSPIGGLIIGLIAGLQVPIVYKIVEKAG 323
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 24, 2003, 22:11:49  
Job time : 6.18991 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.6949 Seconds  
(without alignments)  
2342.388 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVLGYDGP.....PFLLTGLKLYEKNTKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1259	95.3	258	2	Q9F985
2	975	73.8	254	16	Q9KGH5
3	855.5	64.8	259	16	Q8YAC5
4	848.5	64.2	259	16	Q92F54
5	785	59.4	255	16	Q8R7M2
6	767	58.1	259	16	Q8XHL5
7	755	57.2	273	16	Q97EB4
8	663	50.2	265	16	Q9X8N6
9	575	43.5	261	16	Q9A6Z1
10	504.5	38.2	274	16	Q9CD56
11	479.5	36.3	272	16	Q96282
12	415.5	31.5	262	16	Q9RX34
13	409	31.0	256	16	Q8RFE4
14	375.5	28.4	246	16	Q9WZY5
15	352.5	26.7	212	2	Q32514
16	342.5	25.9	273	16	Q83446

17	256.5	19.4	262	16	O51477
18	162	12.3	56	2	P94305
19	158	12.0	295	16	Q8Y2M4
20	150	11.4	257	16	P74045
21	148.5	11.2	229	16	O67753
22	148	11.2	276	16	O8YOD7
23	146	11.1	592	16	O9JXF1
24	146	11.1	592	16	O9JWF1
25	123.5	9.3	248	16	O9HWC1
26	115	8.7	242	16	O9PC14
27	105.5	8.0	597	17	O30225
28	105	7.9	224	16	O8R093
29	103.5	7.8	512	16	O8UBH3
30	99.5	7.5	223	16	O9ZXF6
31	97	7.3	434	16	O97LQ4
32	96	7.3	394	16	O8ZOX3
33	95	7.2	5825	10	O82731
34	94	7.1	148	16	O8627
35	94	7.1	153	16	O92VL2
36	94	7.1	232	16	O92F19
37	94	7.1	520	16	O8UGP6
38	94	7.1	816	16	O8RB00
39	93.5	7.1	465	5	O9U2M4
40	93	7.0	422	16	O8Z41
41	92.5	7.0	212	16	O98D86
42	92	7.0	601	16	O9CPB2
43	91	6.9	394	16	O8ZJ26
44	90	6.8	318	10	O9LVS3
45	90	6.8	405	16	O98AS0

ALIGNMENTS

RESULT 1

Q9F985 PRELIMINARY; PRT; 258 AA.

AC Q9F985; DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative 32 kDa replication protein.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-V;  
RA Vasquez C., Pichuanes S., Saavedra C.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF198621; AAG28531.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRfams; TIGR00671; Baf; 1.  
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match	95.3%;	Score 1259;	DB 2;	Length 258;
Best Local Similarity	94.6%;	Pred. No. 6.8e-107;		
Matches 243;	Conservative 8;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	MIFVLDVGNNTVLGYDDELKHHRIETSRKTEDEYGMKIKALLNHVQLQFSDI	RG I	60
Db	1	MIFVLDVGNNTVLGYDDELKHHRIETSRKTEDEYGMKIKALLNHVQLQFSDI	RG I	60
QY	61	IISSVVPPIFALERCLKYFHIKPLIVGPGIKTGLDIKNDNPREVGADRI	VNAGIHL	120
Db	61	IISSVVPPIFALERCLKYFHIKPLIVGPGIKTGLNIKNDNPREVGADRI	VNAVAGIHL	120
QY	121	YGSPLIIVDFGTATTTCYINEHKQYMGGAITAPGIMISTEALFRAAKLPRI	ETARPDII	180
Db	121	YGSPLIIVDFGTATTTCYINEHKQYMGGAITAPGIMISTEALFRAAKLPRI	ETARPDII	180

```

QY 181 GNTVSMQAGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPPF 240
Db 181 GNTVSMQAGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESNVIDVDPPF 240
QY 241 LTLTGLKLLYEKNTEKK 257
Db 241 LTLTGLKLLYEKNVDKK 257

RESULT 2
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
X NCBI_TaxID=86665;
AN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -.
DR InterPro; IPR004619; Baf..
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 73.8%; Score 975; DB 16; Length 254;
Best Local Similarity 70.8%; Pred. No. 5.8e-81;
Matches 179; Conservative 36; Mismatches 38; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRLETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
Db 1 MILVIDGNTNTVLGVYQDETLYVHHWRLETSRKTEDEYAMTVRSFDHAGLQFQIDGI 60
QY 61 IISVVPPIMFALERMCKYFHKPLIVGPIKTLGIDIKYDNPREGVADRIYNVAGIHL 120
61 VISSVPPMFMFSLQMKCKYFHVTPMIIIGPIKTLGINKYDNPKEVGADRIYNVAAIEL 120
QY 121 YGSLPIIVDFGTATTCYINEKQYMGGAIPGIMISTEALFARAALPRIETARPDDII 180
121 YGPAIVVDGTTATTCYINEKQYAGGVIAPIGIMISTEALYHRAKLPRIETAKPKQVV 180
QY 181 GNTVSMQAGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPPF 240
Db 181 GTNTIDSMQSGIFGYGVQVDGVVVKRMAQAESEPKVIATGGLAKLTGSETIDVIDSF 240
QY 241 LTLTGLKLLYEKN 253
Db 241 LTLKGLQLIYKKN 253

RESULT 3
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.

```

```

OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madoeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -.
DR MEROPS; M41.009; -.
DR ListiList; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 64.8%; Score 855.5; DB 16; Length 259;
Best Local Similarity 62.9%; Pred. No. 4.9e-70;
Matches 163; Conservative 41; Mismatches 52; Indels 3; Gaps 1;

QY 1 MIFVLDVGNNTVLGVYDGDDELKHHWRLETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
Db 1 MILVIDGNTNTVLGVYQDETLYVHHWRLETSRKTEDEYAMTVRSFDHAGLQFQIDGI 60
QY 61 IISVVPPIMFALERMCKYFHKPLIVGPIKTLGIDIKYDNPREGVADRIYNVAGIHL 120
61 VISSVPPIMHAMETCMCVRYFNIRPLIVGPIKTLGINKYDNPREGVADRIYNVAASEE 120
QY 121 YGSLPIIVDFGTATTCYINEKQYMGGAIPGIMISTEALFARAALPRIETARPDDII 180
121 YGTPVIVDGTATTCYIDEGYVQGGAIAPGIMISTEALYHRAKLPRIETARVIAESSQII 180
QY 181 GNTVSMQAGILYGVQVEGIVSRMKAKSKIPPVKVIATGGLAPLIASESDIIDVDPPF 240
Db 181 GKSTVSSMQAGIFGVQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVDILDPF. 240
QY 241 LTLTGLKLLYEKN---TEK 256
241 LTLKGLLELYRRNKPTTEK 259

RESULT 4
Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

```



	Best Local Similarity	56.5%	Pred.	No. 1.3e-63;	Gaps	0
Matches	144;	Conservative	52;	Mismatches	59;	Indels
Qy	1 MIFVLDVGNTNVLGVYDGDDELKHHWRIETSRSTEDGYGMIKALLNHVGLQPSDIRGI	60				
Dd	1 MLAFDVGNVTNVWGVPFGKKLLHSFRISTDKNKTYDEYGMIVNLGIGINGISLTFEIDDDV	60				
Qy	61 IISVVPPIMFALERMCLKFYHIKPLIYGPGIKTGDIKYDNPREVGADRINNAVAGIH L	120				
Dd	61 IISVVPPLMNTLVQMSLKFVFRTKPVI VVGPIKTGINIKYDNPREVGADRVNAVAAYEL	120				
Qy	121 YGSPLIIDPGTATTTCYYINEHKYOYMGATAPGMTIMEALFAAAKLPRIEIARPDII	180				
Dd	121 YGGPVIVIDGTATTTCACISEKGBYLGIIITAPGLMISADALFORTAKLPKDILT KPPTVI	180				
Qy	181 GKNYSMAQGAILGYGVQVEGISVRMKASKIPPKVITATCGTCLAPLIASESDIIDVDPF	240				
Dd	181 NNRTVASMQSGILYGHVCMWDYIVTRMKGEAPSAYVAVTGCGFANMIAESEKT IDTVNM	240				
Qy	241 LTLTGLKLLYEKNTE	255				
Dd	241 LTLEGLRIIYERNKE	255				
RESULT 6	ID Q8XHL5 PRELIMINARY; PRT; 259 AA.					
AC	Q8XHL5;					
DT	01-MAR-2002 (TrEMBLrel. 20, Created)					
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)					
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)					
DE	Hypothetical protein CPE2468.					
OS	Clostridium perfringens.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;					
OX	Clostridiales; Clostridiaceae; Clostridium.					
NCBI_TaxId=1502;	[1]					
RN	SEQUENCE FROM N.A.					
RC	STRAIN=13 / TYPE A;					
RX	PubMed=11792842;					
RA	Shinizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,					
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;					
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic					
RL	flesh-eater." ;					
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).					
DR	ENBL: AP003194; BAB82174.1; -					
DR	InterPro: IPR004619; Baf.					
DR	InterPro: IPR000515; BPD_transp.					
DR	Pfam: PF03309; Bvg_acc_factor; 1.					
DR	TIGRFAMS; TIGR00671; baf; 1.					
DR	PROSITE; PS00402; BPD_TRANS_PNN_MEMBR; UNKNOWN_1.					
KQ	Hypothetical protein; Complete proteome.					
SQ	SEQUENCE 259 AA; 28619 MW; OD5FEA3B7A145E10 CRC64;					
Query Match	58.1%; Score 767; DB 16; Length 259;					
Best Local Similarity	58.7%;	Pred.	No. 6e-62;			
Matches	152;	Conservative	37;	Mismatches	66;	Indels
Qy	1 MIFVLDVGNTNVLGVYDGDDELKHHWRIETSRSTEDGYGMIKALLNHVGLQPSDIRGI	60				
Dd	1 MILLIDVGNTNVILGIHDNEKIYASRWISTDSKTSDEYSTSQVMQLFNQAALNPEDVEGI	60				
Qy	61 IISVVPPIMFALERMCLKFYHIKPLIYGPGIKTGDIKYDNPREVGADRINNAVAGIH L	120				
Dd	61 IISVVPNIMHLSLENMVVKCFCKEPVI VVGPIKTGINIKYDNPREVGADRVNAVAAPEK	120				
Qy	121 YGSPLIIDPGTATTTCYYINEHKYOYMGATAPGMTIMEALFAAAKLPRIEIARPDII	180				
Dd	121 HKRPMLIIDGTATTTCATYETKDYLGNGICPGIQISNDALFERAAKLPRILELKPSVI	180				
Qy	181 GKNTVSMAQGAILGYGVQVEGISVRMKAK-----SKIPPKVYTAGGPLIASESDIIDV	236				

```

Db 181 CKNFTVSMQAGIIGYGVIGKVEIVKRMKKEMDGLGEKPEFVLATGGLAKLIVSETVIDE 240
QY 237 VDPFLLTGLKLLYEKNTE 255
db 241 VDRKLTLEGLKIYEKNKE 259

RESULT 7
Q97EB4 PRELIMINARY; PRT; 273 AA.
ID Q97EB4;
AC Q97EB4;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
NCBI_TaxID=1488;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.I., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838 (2001).
DR EMBL; AF007815; AAK81136.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 57.2%; Score 755; DB 16; Length 273;
Best Local Similarity 56.2%; Pred. No. 8e-61;
Matches 146; Conservative 47; Mismatches 63; Indels 4; Gaps 1;

QY 1 MIFVLDVGNNTVNLGVYDGDGLKHHWRIETSRKSTEDYGMKIKALLNHVGLQFSDIRGI 60
Db 12 VILVLDVGNNTVNLGVYDNDTKLTAEWRLSTDVLRNDEYGIQVMNLFQODKLDPLVEGV 71
61 IISVVPPIPMFALERMKLYFIHKPLIVGPGIKTGLDIDKYDNPREGVADRIYNVAGIHL 120
Db 72 IISVVPNIMVLSLHMRIRKRYINPLVGVPGIKTGINIKYDNPKEVGADRIYNVAAHEI 131
QY 121 VGSPLIIVDFGTATTCYINHEKQYMGGAIAPIGMISTEALFARAAPRIAPDDII 180
Db 132 YARSIIIDFGATTTCVAVRENGDYLGGAICPGIKVSSEALFARAAPRIAPRIAPAYAI 191
QY 181 GKNTVSAMQAGIIGYGVIGKVEIVKRMKKEMDGLGEKPEFVLATGGLAKLISEAKNDV 236
Db 192 CKNTISSQSIGVYIGQVRYIVRMKEELQEKEKEPELVVATGGLAKLISEAKNDV 251
QY 237 VDPFLLTGLKLLYEKNTEK 256
Db 252 INPFLTLEGLRIYEKNVK 271

RESULT 8
Q9X8N6 PRELIMINARY; PRT; 265 AA.
ID Q9X8N6
AC Q9X8N6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

```

```

DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.3IC.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bacteman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147 (2002).
DR EMBL; AL049628; CAB40880.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 50.2%; Score 663; DB 16; Length 265;
Best Local Similarity 48.9%; Pred. No. 1.9e-52;
Matches 129; Conservative 58; Mismatches 69; Indels 8; Gaps 5;

QY 1 MIFVLDVGNNTVNLGVYDGDGLKHHWRIETSRKSTEDYGMKIKALLN-H-VGLQFSD- 56
Db 1 MLTIDVGNTHVNLGVLDGEDIVHWRISTDSRTADELAVLQGLMGHPLGLDELGDG 60
QY 57 IRGIISSVVPPIPMFALERMKLYFIHKPLIVGPGIKTGLDIDKYDNPREGVADRIYNV 115
Db 61 IDGIAICATVPVSLHMRIRKRYINPLVGVPGIKTGINIKYDNPKEVGADRIYNV 120
QY 116 AGIHLVGSPLIIVDFGTATTCYINHEKQYMGGAIAPIGMISTEALFARAAPRIEIR 175
Db 121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPIGIEISVEALGVGAQLKIEVAR 180
QY 176 PDDIIGKNTVSAMQAGIIGYGVIGKVEIVKRMKKEMDGLGEKPEFVLATGGLAKLISE 232
Db 181 PRSVIGKNTVSAMQAGIIGYGVIGKVEIVKRMKKEMDGLGEKPEFVLATGGLAKLISE 240
QY 233 IIDVVDPFLLTGLKLLYEKNTEK 256
Db 241 VIDEHEPWLTLGLRLVYERNVSR 264

RESULT 9

```

```
Q9A6Z1
ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.
AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=2113698; PubMed=11259647;
RA Eisen J., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Nelson W.C., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714E95 CRC64;

Query Match 43.5%; Score 575; DB 16; Length 261;
Best Local Similarity 44.1%; Pred. No. 2.1e-44;
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;

QY 1 MIFVLDGNTNTVLGVYDGDGLKHHRIETSRKTEDEYGMKALNHVGLQSDIRGI 60
DB 2 MLLAEIGNTNTFAIHGASWVAQWSATESTRTADYVYVWLSQGLSMOGLGFRAIDAV 61
QY 61 IISVWPPIFALERMCLKFKHKLIVGPGIKTGLDTPKYNPREGADRVNVAAGIHL 120
DB 62 IISVWPQSIENLRLSRFVNEPLVIGENAKGIDVRICKPEAGADRVNVAIGAMV 121
QY 121 YGSPLLIVDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRIEIAIP--DD 178
DB 122 YGSPVLDGSGTATTFDVAADGAFEGGIAPGINSQALHEAAAKLPRIAIQPNAGR 181
QY 179 IIGKNTYSAMQAGILYGVGVGEGIVSRMKAKSIPKVIATGGGLPIASESDIIDVVD 238
DB 182 IGVGDTYSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASFEGATSDIHFD 241
QY 239 PFLTLTGLKLYEKNT 254
DB 242 SDLTIRGLLEIYRNT 257

RESULT 10
Q9CD56
ID Q9CD56 PRELIMINARY; PRT; 274 AA.
AC Q9CD56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.

Q9A6Z1
ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.
AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=2113698; PubMed=11259647;
RA Eisen J., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Nelson W.C., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714E95 CRC64;

Query Match 43.5%; Score 575; DB 16; Length 261;
Best Local Similarity 44.1%; Pred. No. 2.1e-44;
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;

QY 1 MIFVLDGNTNTVLGVYDGDGLKHHRIETSRKTEDEYGMKALNHVGLQSDIRGI 60
DB 2 MLLAEIGNTNTFAIHGASWVAQWSATESTRTADYVYVWLSQGLSMOGLGFRAIDAV 61
QY 61 IISVWPPIFALERMCLKFKHKLIVGPGIKTGLDTPKYNPREGADRVNVAAGIHL 120
DB 62 IISVWPQSIENLRLSRFVNEPLVIGENAKGIDVRICKPEAGADRVNVAIGAMV 121
QY 121 YGSPLLIVDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRIEIAIP--DD 178
DB 122 YGSPVLDGSGTATTFDVAADGAFEGGIAPGINSQALHEAAAKLPRIAIQPNAGR 181
QY 179 IIGKNTYSAMQAGILYGVGVGEGIVSRMKAKSIPKVIATGGGLPIASESDIIDVVD 238
DB 182 IGVGDTYSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASFEGATSDIHFD 241
QY 239 PFLTLTGLKLYEKNT 254
DB 242 SDLTIRGLLEIYRNT 257

RESULT 10
Q9CD56
ID Q9CD56 PRELIMINARY; PRT; 274 AA.
AC Q9CD56;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeler K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583917; CAC29740.1; -.
DR Leproma; ML0232; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 38.2%; Score 504.5; DB 16; Length 274;
Best Local Similarity 38.7%; Pred. No. 6.1e-38;
Matches 105; Conservative 58; Mismatches 85; Indels 23; Gaps 6;

QY 1 MIFVLDGNTNTVLGVYDGDGLKHHRIETSRKTEDEYGMKALNHVGLQF 54
DB 1 MLLAIDVRNTHTVGLLSGS--KEHAKVVOQNRIRTESEVTADELALIDGL---IGDSS 55
QY 55 SDIRGIIISVYPPPIWFALERMCLKFKHKLIVGPGIKTGLDTPKYNPREGADRVN 113
DB 56 ERLAGAAALSTVPSVLHEVRIMLDQYWPSPVHVLPEPVRTGIPGLVDNPKEVGADRVN 115
QY 114 AVAGHLXGSPLLIIVDFGTATTCYINEHKQYMGGAIPGIMISTEALFARAALPRIE 173
DB 116 CLAAFKKGQAIVVDVFGSSICVDSVSAKGEFLGGAIAFGVQVSSDAAAARSALRVEL 175
QY 174 ARPDDIIGKNTYSAMQAGILYGVGVGEGIVSRMKAK-----SKIPPKVIATGG 224
DB 176 APRSVVGKNTVECMQAGVFWGAGLVGRLDGLVGRMDVEEFGSDLGNRV--AVVATGHTA 233
QY 225 PLIASESDIIDVVDPLTLTGLKLYEKNT 255
DB 234 PLLLPETHVDHYDRHLTLHLGLRVFERNR 264

RESULT 11
Q06282
ID Q06282 PRELIMINARY; PRT; 272 AA.
AC Q06282;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative, Baf family).
GN RV3600C OR MTCY07H.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
```

```

RT complete genome sequence. *;
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z5557; CAB08944.1; -.
DR EMBL; AE007170; AAK48063.1; -.
DR TIGR; MT3706; -.
DR TubercuList; RV3600c; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 272 AA; 29304 MW; 5D70E6EF0F09AC8B CRC64;

Query Match 36.3%; Score 479.5; DB 16; Length 272;
Best Local Similarity 37.8%; Pred. No. 1.2e-35;
Matches 101; Conservative 56; Mismatches 93; Indels 17; Gaps 5;

QY 1 MIFVLDVGNNTVILGVYDGDDELKHH-----WRIETSRKTEDEYGMKIKALLNHVGLQF 54
DB 1 MLLAIDVNTHTVPGVLLSS--MKEHAKVVQWRIRTESEVTADALALIDGL---IGEDS 55
QY 55 SDIRGIISVVPPIMFALERMLKLYFIKHP-LIVGPGIKTGLDIKYDNPREVGADRIYN 113
DB 56 ERLGTGTAALSTVPSVLHEVRIMLDQWPSVHLIEPVGRTGIPILLVDNPKVEGADRIYN 115
QY 114 AVAGIHLVGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAAKLPRIEI 173
DB 116 CLAAVDPRKAAIIVDFGSSICVDVWSAKGFLGGAIAAGVQVSSDAARSAALRRVEL 175
QY 174 ARPDDIIGKNTVSAMQAGILGYGVQVEGIYSRMAK-----KSKIPPKVIATGGIAPLIA 228
DB 176 APRSRVGNKNTVECMQAGVAGVAGLVDGLVGRIDREYSGFSDVDHDAIVATGHTAPLL 235
QY 229 SESDIIDVDFPLITGLKLYEKNT 255
DB 236 PELHTVDHYDQHLTQLGLRLVERNLE 262

RESULT 12
3RX54
.D Q9RX54 PRELIMINARY; PRT; 262 AA.
AC Q9RX54;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RA MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001905; AAF10040.1; -.
DR TIGR; DR0461; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
W Hypothetical protein; Complete proteome.
Q SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 31.5%; Score 415.5; DB 16; Length 262;
Best Local Similarity 35.2%; Pred. No. 7.8e-30;
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;

QY 2 IFVLDVGNNTVILGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRGI 60
DB 6 LLAVDIGNTTVLGLADASGALHTWIRINREMLPDDLALQJLHGLTLAGAIPIP--RAA 63
QY 61 IISVVVPPI-----MFALERMLKLYFIKHPILVPGIKTGLDIKYDNPREVGADRIYN 116
DB 64 VLSVAPPVGENYALAKR-----HFMDAFVSAENLPDVTVELDTPGSGVADRLCN--- 116
QY 117 GIHLVGP-----LIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARA 168
DB 117 ---LFGAEKYLGLDYAVVDFGTSTNFDVGRGRFLGILATGAQVSADALFARA 173
QY 169 PRIETARPDIDIGKNTVSAMQAGILGYGVQVEGIYSRMAKSKIPPKVIATGGIAPLIA 228
DB 174 PRITLOAPETAIGKNTVHALQSLVFGYAEVMDGLLRIRRAELPGEAVAVATGGFSRT 233
QY 229 SESDIIDVDFPLITGLKLYEKNT 255
DB 234 GICQEIIDYDETLTGLGLVELNARS 260

RESULT 13
Q8RFE4
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RA MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Rzenik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010586; AAL94957.1; -.
W Complete proteome.
Q SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 31.0%; Score 409; DB 16; Length 256;
Best Local Similarity 37.5%; Pred. No. 3e-29;
Matches 96; Conservative 46; Mismatches 110; Indels 4; Gaps 4;

QY 1 MIFVLDVGNNTVILGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRG 59
DB 1 MIIGDIGNTHIVGIYDNGELISTFRIATNDKMTDEDEYFYFNNTKYNEISIKKVD 60

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.4678 Seconds  
(without alignments)  
725.198 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVVLGVYDGD.....PFLTLTGLKLLYKNTKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Issued Patents AA.\*
- 1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pap.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pap.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pap.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pap.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pap.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	7.0	330	4	US-09-134-001C-3582
2	84	6.4	342	4	US-08-096-181A-12
3	84	6.4	342	4	US-08-096-181A-14
4	84	6.4	342	5	PCT-US94-08326-12
5	84	6.4	342	5	PCT-US94-08326-14
6	84	6.4	361	4	US-08-096-181A-8
7	84	6.4	361	5	PCT-US94-08326-8
8	84	6.4	363	4	US-08-096-181A-10
9	84	6.4	363	5	PCT-US94-08326-10
10	83	6.3	365	1	US-08-447-500-4
11	83	6.3	365	1	US-08-454-097-4
12	83	6.3	365	1	US-08-453-866-4
13	83	6.3	365	3	US-09-100-664A-13
14	83	6.3	365	3	US-08-185-359-4
15	83	6.3	365	4	US-09-335-983-13
16	82	6.2	314	4	US-09-134-001C-2983
17	81.5	6.2	494	1	US-08-447-500-2
18	81.5	6.2	494	1	US-08-454-097-2
19	81.5	6.2	494	1	US-08-447-408-2
20	81.5	6.2	494	1	US-08-453-866-2
21	81.5	6.2	494	3	US-08-185-359-2
22	81.5	6.2	494	4	US-09-457-040B-25
23	79.5	6.0	421	4	US-09-134-001C-5201
24	79	6.0	489	4	US-09-431-099-2
25	79	6.0	489	4	US-09-431-099-4
26	78.5	5.9	1012	2	US-08-475-891A-4
27	78.5	5.9	1025	2	US-08-567-375-4

28	78.5	5.9	1025	2	US-08-587-680A-4	Sequence 4, Appli
29	76.5	5.8	201	2	US-08-911-364-2	Sequence 2, Appli
30	76.5	5.8	733	2	US-08-911-364-1	Sequence 1, Appli
31	76.5	5.8	733	4	US-08-464-700-2	Sequence 2, Appli
32	76	5.8	248	4	US-09-134-001C-3211	Sequence 3211, Ap
33	76	5.8	399	4	US-09-134-001C-4571	Sequence 4571, Ap
34	75.5	5.7	542	3	US-08-968-563-16	Sequence 16, Appl
35	75.5	5.7	542	4	US-08-969-683A-16	Sequence 16, Appl
36	75.5	5.7	542	4	US-09-297-928-12	Sequence 12, Appl
37	74.5	5.6	609	4	US-08-969-683A-67	Sequence 67, Appl
38	74.5	5.6	1375	4	US-09-171-410-1	Sequence 1, Appli
39	73.5	5.6	340	4	US-09-789-300A-2	Sequence 2, Appli
40	73.5	5.6	625	1	US-08-391-615-3	Sequence 3, Appli
41	73	5.5	724	3	US-08-793-331-4	Sequence 4, Appli
42	73	5.5	725	3	US-08-793-331-6	Sequence 6, Appli
43	73	5.5	2052	3	US-09-045-201A-2	Sequence 2, Appli
44	73	5.5	2052	4	US-09-619-062-2	Sequence 2, Appli
45	72.5	5.5	521	1	US-08-915-003-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-09-134-001C-3582  
; Sequence 3582, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3582  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3582

Query Match	7.0%;	Score 92;	DB 4;	Length 330;
Best Local Similarity	20.6%;	Pred. No. 0.014;		
Matches	60;	Conservative 40;	Mismatches 87;	Indels 104; Gaps 16;
QY	1	MIFVLDVGNNTVVLGVYDGD-ELKHHWRIETSRKTEDEYGMKIKALLNHVGLQFSDIRG	59	
Db	6	IILAADIGGTTCKLGIFDKDLQLHKWSIDTD---TSDHTG---ELLKNYNSFT----	55	
QY	60	IIISVVPIPMFALERMCLYFKPIKPLIVGPGIKTGLDIKYDNPREFGADRIYNVAGIH	119	
Db	56	-----BKIA-----EYKIDFNNVVG-----VGI 74		
QY	120	LYSGPLIIVDFGTAATTCYINEH-----KOYMGGAIPAGIMISTEALFAAKL	168	
Db	75	VSPG---VDFTGTVYGVAVNLHPDVSNNVREIFKQYVN---CP-VYVNDANVAALGEK	126	
QY	169	PRIETARPDIIIGKNTVSAMQAGIL-----YGYVQVGVGVSRMKA-----KS	211	
Db	127	HKGAGGADVVVAITLGTGLGGIISNGEIVHGHNGS-GAEIGHLRADQRFQCNCGKS	185	
QY	212	KTPPKVIATGGLAPLIASESDIIDVD---PFLTLG--LKLLYKNTKK 257		
Db	186	-----GCETVASATGVNVLNFFYYPKLTFFKKSILQLIKDNOVTAK	226	

RESULT 2  
US-08-096-181A-12  
; Sequence 12, Application US/08096181A

```
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/096,181A
; FILING DATE: 23-Jul-1993
; CLASSIFICATION: 424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-181A-12

Query Match 6.4%; Score 84; DB 4; Length 342;
Best Local Similarity 23.0%; Pred. No. 0.13;
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73
DB 100 VKLGRKTIADGITSADKEYGLNNSDYIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156
QY 74 ERMCLKYPHIKPLIVPGIKTGLDIKYNPREVGADRIYNAVAGIHLGSPLLIYDFGTA 133
DB 157 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 203
QY 134 TTYCY--INEHKQYMGGAIA-----GIMISTEALFAR 164
DB 204 TNYKYNESDEHKQQLNGVLATIGYRFSDLGLLVSLDSGYAK 244

RESULT 3
US-08-096-181A-14
; Sequence 14, Application US/08096181A
; Patent No. 6153406
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/096,181A
; FILING DATE: 23-Jul-1993
; CLASSIFICATION: 424
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-096-181A-14

Query Match 6.4%; Score 84; DB 4; Length 342;
Best Local Similarity 23.0%; Pred. No. 0.13;
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73
DB 100 VKLGRKTIADGITSADKEYGLNNSDYIPTSGNTVGYTFKGDGLVGA---NYLLAQ 156
QY 74 ERMCLKYPHIKPLIVPGIKTGLDIKYNPREVGADRIYNAVAGIHLGSPLLIYDFGTA 133
DB 157 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 203
QY 134 TTYCY--INEHKQYMGGAIA-----GIMISTEALFAR 164
DB 204 TNYKYNESDEHKQQLNGVLATIGYRFSDLGLLVSLDSGYAK 244
```

```
RESULT 4
PCT-US94-08326-12
; Sequence 12, Application PC/TUS9408326
; GENERAL INFORMATION:
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
; TITLE OF INVENTION: A Method For The High Level
; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane
; TITLE OF INVENTION: Protein
; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: PCT/US94/08326
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,181
; FILING DATE: 23-JULY-1993
```



ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REFERENCE/DOCKET NUMBER: 1438.001PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-08326-12

Query Match 6.4%; Score 84; DB 5; Length 342;  
Best Local Similarity 23.0%; Pred. No. 0.13;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
DB 100 VKLGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKIDGLVIGA---NYLLAQ 156  
QY 74 ERMCLKYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAGIHLVYGSPLIIIVDFGTA 133  
DB 157 KREGAKGENKRPNDKAGEVRIG---EINNGIOVGAKYDANDIVAKIAYG-----R 203  
QY 134 TTYCY--INEHKQYMGGAIA-----GIMISTEALFAR 164  
DB 204 TNYKYNESDEHKQOLNGVLATIGYRFSDLGLVSLDSGYAK 244

RESULT 5  
PCT-US94-08326-14  
Sequence 14, Application PC/TUS9408326  
GENERAL INFORMATION:  
APPLICANT: North American Vaccine, Inc.  
APPLICANT: 12103 Indian Creek Court  
APPLICANT: Beltsville, MD 20705  
APPLICANT: Pullen, Jeffrey K.  
APPLICANT: Soper, Thomas S.  
APPLICANT: Liang, Shu-Mei  
TITLE OF INVENTION: A Method For The High Level  
TITLE OF INVENTION: Expression,  
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane  
TITLE OF INVENTION: Protein  
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/08326  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/096,181  
FILING DATE: 23-JULY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REFERENCE/DOCKET NUMBER: 1438.001PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-08326-14

Query Match 6.4%; Score 84; DB 5; Length 342;  
Best Local Similarity 23.0%; Pred. No. 0.13;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
DB 100 VKLGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKIDGLVIGA---NYLLAQ 156  
QY 74 ERMCLKYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAGIHLVYGSPLIIIVDFGTA 133  
DB 157 KREGAKGENKRPNDKAGEVRIG---EINNGIOVGAKYDANDIVAKIAYG-----R 203  
QY 134 TTYCY--INEHKQYMGGAIA-----GIMISTEALFAR 164  
DB 204 TNYKYNESDEHKQOLNGVLATIGYRFSDLGLVSLDSGYAK 244

RESULT 6  
US-08-096-181A-8  
Sequence 8, Application US/08096181A  
Patent No. 6153406  
GENERAL INFORMATION:  
APPLICANT: Tai, Joseph Y.  
APPLICANT: Pullen, Jeffrey K.  
APPLICANT: Soper, Thomas S.  
APPLICANT: Liang, Shu-Mei  
TITLE OF INVENTION: A Method For The High Level Expression,  
TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein  
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,181A  
FILING DATE: 23-Jul-1993  
CLASSIFICATION: 424  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 361 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-096-181A-8  
Query Match 6.4%; Score 84; DB 4; Length 361;  
Best Local Similarity 23.0%; Pred. No. 0.15;  
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;  
QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73  
DB 119 VKLGRAKTIADGITSADKEYGVLNNSDIPTSGNTVGYTFKIDGLVIGA---NYLLAQ 175  
QY 74 ERMCLKYFHKPLIVGPGIKTGLDIKYNPREVGADRIVNAGIHLVYGSPLIIIVDFGTA 133

Db 176 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 222

QY 134 TTYCY--INEHKQYMGGAIP-----GIMISTEALFAR 164

Db 223 TNYKNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAK 263

## RESULT 7

PCT-US94-08326-8

; Sequence 8, Application PC/TUS9408326

; GENERAL INFORMATION:

; APPLICANT: North American Vaccine, Inc.

; APPLICANT: 12103 Indian Creek Court

; APPLICANT: Beltsville, MD 20705

; APPLICANT: Pullen, Jeffrey K.

; APPLICANT: Soper, Thomas S.

; APPLICANT: Liang, Shu-Mei

; TITLE OF INVENTION: A Method For The High Level

; TITLE OF INVENTION: Expression,

; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane

; TITLE OF INVENTION: Protein

; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/08326

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/096,181

; FILING DATE: 23-JULY-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REFERENCE/DOCKET NUMBER: 1438.001PC01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US94-08326-8

## Query Match

Best Local Similarity 6.4%; Score 84; DB 5; Length 361;

Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73

Db 119 VKLGRAKTIADGITSADKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVGA---NYLLAQ 175

QY 74 ERMLKYFHKPLIVGPIKTKLDIKYDNPREVGADRVNAVAGIHLVGSPLIIVDFGTA 133

Db 176 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 222

QY 134 TTYCY--INEHKQYMGGAIP-----GIMISTEALFAR 164

Db 223 TNYKNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAK 263

## RESULT 8

US-08-096-181A-10

; Sequence 10, Application US/08096181A

; Patent No. 6153406

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Pullen, Jeffrey K.

; APPLICANT: Soper, Thomas S.

; APPLICANT: Liang, Shu-Mei

; TITLE OF INVENTION: A Method For The High Level Expression,

; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein

; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein &amp; Fox

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/096,181A

; FILING DATE: 23-Jul-1993

; CLASSIFICATION: 424

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-096-181A-10

## Query Match

6.4%; Score 84; DB 4; Length 363;

Best Local Similarity 23.0%; Pred. No. 0.15;

Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSRKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPPIMFAL 73

Db 121 VKLGRAKTIADGITSADKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVGA---NYLLAQ 177

QY 74 ERMLKYFHKPLIVGPIKTKLDIKYDNPREVGADRVNAVAGIHLVGSPLIIVDFGTA 133

Db 178 KREGAKGNKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYG-----R 224

QY 134 TTYCY--INEHKQYMGGAIP-----GIMISTEALFAR 164

Db 225 TNYKNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAK 265

## RESULT 9

PCT-US94-08326-10

; Sequence 10, Application PC/TUS9408326

; GENERAL INFORMATION:

; APPLICANT: North American Vaccine, Inc.

; APPLICANT: 12103 Indian Creek Court

; APPLICANT: Beltsville, MD 20705

; APPLICANT: Pullen, Jeffrey K.

; APPLICANT: Soper, Thomas S.

; APPLICANT: Liang, Shu-Mei

; TITLE OF INVENTION: A Method For The High Level

; TITLE OF INVENTION: Expression,

; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane

; TITLE OF INVENTION: Protein

; TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

```

; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08326
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,181
; FILING DATE: 23-JULY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 1438.001PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08326-10

Query Match 6.4%; Score 84; DB 5; Length 363;
Best Local Similarity 23.0%; Pred. No. 0.15;
Matches 37; Conservative 30; Mismatches 54; Indels 40; Gaps 7;

QY 28 IETSKT-----EDEYGM-----IKALLNHVGLQFSDIRGIISVVPIMFAL 73
DB 121 VRLGRAKTIADGITSADKEYGLVNNSDYIPTSGNTVGYTFKGDGLVGA---NYLLAQ 177
QY 74 ERMCLKYPIIKLIVPGIKTGLDIKYNPREVGADRIYNVAGIHLGSPLIIVDFGTA 133
DB 178 KREGAKGENKPNDRKAGEVRIG---EINNGIQVGAKYDANDIVAKTAYG-----R 224
QY 134 TTYCY--INEHKQYMGGAIP-----GIMISTEALFAR 164
DB 225 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLVSLDSGYAK 265

RESULT 10
US-08-447-500-4
; Sequence 4, Application US/08447500
; Patent No. 5627064
; GENERAL INFORMATION:
; APPLICANT: Hoeckstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,500
```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/008,001
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-2458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-500-4

Query Match 6.3%; Score 83; DB 1; Length 365;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 51; Conservative 47; Mismatches 92; Indels 32; Gaps 11;

QY 16 VYDGLKHHWRIETSRK---TEDEYGMKALLNHVGLQFSDIRGI-----IISVV 66
DB 31 VVSGEEVA--IKLESTRAKHPQLEYEY-RVYRILSGVGIPFVRWFVGECDYNAMVMDLL 87
QY 67 PTIMFALERMCLKYPIIKP-LIVPGIKTGLDIKYNP---REVGADRIYNVAGIHLG 122
DB 88 GPSLEDLFNFCNRKFSKLVLLADQLISRIEHSKSFHLRDKPD---NFLMGIGKRG 144
QY 123 SPLIIVDFGTATTCYINNEHKQYMGGAIPGIMISTEALFARAALPRIETARDDIIGK 182
DB 145 NOVNIIDFGLAKKY---RDHKTHLHIPYRENKNTLGTARYASINTHLGIEQSRDDLES 201
QY 183 NTVSAMQAGILGYGVQV--EGIVSRMKAK--SKIPPKVIAT 220
DB 202 GYV-----LVYFCRGLPWOGLKATTKKOKYKIMEKKIST 237

RESULT 11
US-08-454-037-4
; Sequence 4, Application US/08454097
; Patent No. 5686412
; GENERAL INFORMATION:
; APPLICANT: Hoeckstra, Merl F.
; TITLE OF INVENTION: Protein Kinases
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,097
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/185,359
; FILING DATE: 21-JAN-1994
; APPLICATION NUMBER: US 08/008,001
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,783
; FILING DATE: 03-JUL-1991
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5686412and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31853
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-454-097-4

Query Match      6.3%; Score 83; DB 1; Length 365;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 51; Conservative 47; Mismatches 92; Indels 32; Gaps 11;

Y       16 VYDDELKHHWRIETSRSK---TEDEYGMIMKALLNHVGLQSDIRGI-----IISVV 66
Db      31 VVSSEVA--IKLESTRAKHPOLEYEY-RVYRILSGVGIPFWFGVECDYNAMYMDLL 87
QY      67 PPIWFALERMCLKYFHKP-LIVPGIGKTGLDIKYDNP---REVGAADRVNAVAGHLYG 122
Db      88 GPSLEDLFPNCRKFSLKTVLLADOLISRIFHSKFLHRDIKPD---NFLMGICKRG 144
QY     123 SPLIIVDFGTATTCYINHEKHQYMGAIAPIGMISTEALFAARAUKLPRIEIARPPDDIIGK 182
Db     145 NQVNIIDFDGLAKKY---RDHKTHLHPYRENKNLTGTARYASINTHLGIEQSRDDLES 201
QY     183 NTYSMAQAGILGYGVQV---EGIVSRMKAK--SKIPPKVIAT 220
Db     202 GYV-----LVYFCRGLPWOGLKATTKKKYEKIMEKKIST 237

RESULT 12
US-08-453-866-4
; Sequence 4, Application US/08453866
; Patent No. 5756289
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Juba & Lubitz
; STREET: 1890 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,866
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/008,001
; FILING DATE: 20-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-2458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-100-664A-13

; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-453-866-4

Query Match      6.3%; Score 83; DB 1; Length 365;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 51; Conservative 47; Mismatches 92; Indels 32; Gaps 11;

QY      16 VYDDELKHHWRIETSRSK---TEDEYGMIMKALLNHVGLQSDIRGI-----IISVV 66
Db      31 VVSSEVA--IKLESTRAKHPOLEYEY-RVYRILSGVGIPFWFGVECDYNAMYMDLL 87
QY      67 PPIWFALERMCLKYFHKP-LIVPGIGKTGLDIKYDNP---REVGAADRVNAVAGHLYG 122
Db      88 GPSLEDLFPNCRKFSLKTVLLADOLISRIFHSKFLHRDIKPD---NFLMGICKRG 144
QY     123 SPLIIVDFGTATTCYINHEKHQYMGAIAPIGMISTEALFAARAUKLPRIEIARPPDDIIGK 182
Db     145 NQVNIIDFDGLAKKY---RDHKTHLHPYRENKNLTGTARYASINTHLGIEQSRDDLES 201
QY     183 NTYSMAQAGILGYGVQV---EGIVSRMKAK--SKIPPKVIAT 220
Db     202 GYV-----LVYFCRGLPWOGLKATTKKKYEKIMEKKIST 237

RESULT 13
US-09-100-664A-13
; Sequence 13, Application US/09100664A
; Patent No. 6057129
; GENERAL INFORMATION:
; APPLICANT: YOUNG, MICHAEL W.
; APPLICANT: KLOSS, BRIAN
; APPLICANT: BLAU, JUSTIN
; APPLICANT: PRICE, JEFFREY
; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,664A
; FILING DATE: 19-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-100-664A-13

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.3382 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-49  
Perfect score: 1321  
Sequence: 1 MIFVLDVGNNTVLGVYDGD.....PFLTLTGKLLYEKNTKKG 258

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PC0\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PC0US\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1321	100.0	258	9	US-09-813-453A-49
2	1057.5	80.1	262	9	US-09-813-453A-45
3	1046	79.2	258	9	US-09-813-453A-2
4	975	73.8	254	9	US-09-813-453A-47
5	900	68.1	233	9	US-09-813-453A-17
6	834	63.1	255	9	US-09-813-453A-7
7	817.5	61.9	256	9	US-09-813-453A-55
8	663	50.2	265	9	US-09-813-453A-4
9	629.5	47.7	250	9	US-09-813-453A-3
10	575	43.5	260	9	US-09-813-453A-51
11	542	41.0	258	9	US-09-813-453A-6
12	513	38.8	219	9	US-09-813-453A-57
13	479.5	36.3	272	9	US-09-813-453A-5
14	475.5	36.0	272	9	US-09-712-363-276
15	415.5	31.5	262	9	US-09-813-453A-8
16	375.5	28.4	246	9	US-09-813-453A-9
17	352.5	26.7	212	9	US-09-813-453A-59
18	342.5	25.9	273	9	US-09-813-453A-10
19	305	23.1	257	9	US-09-813-453A-53

20	256.5	19.4	262	9	US-09-813-453A-11	Sequence 11, Appl
21	210.5	15.9	244	9	US-09-813-453A-41	Sequence 41, Appl
22	178.5	13.5	241	9	US-09-813-453A-63	Sequence 63, Appl
23	157	11.9	249	9	US-09-813-453A-70	Sequence 70, Appl
24	150	11.4	257	9	US-09-813-453A-13	Sequence 13, Appl
25	148.5	11.2	229	9	US-09-813-453A-12	Sequence 12, Appl
26	146	11.1	592	9	US-09-813-453A-22	Sequence 22, Appl
27	146	11.1	592	9	US-09-813-453A-43	Sequence 43, Appl
28	145	11.0	460	9	US-09-813-453A-39	Sequence 39, Appl
29	140.5	10.6	267	9	US-09-813-453A-15	Sequence 15, Appl
30	139	10.5	249	9	US-09-813-453A-61	Sequence 61, Appl
31	123.5	9.3	248	9	US-09-813-453A-20	Sequence 20, Appl
32	115	8.7	242	9	US-09-813-453A-65	Sequence 65, Appl
33	87.5	6.6	223	9	US-09-895-913A-74	Sequence 74, Appl
34	87.5	6.6	223	9	US-09-813-453A-4	Sequence 4, Appl
35	87.5	6.6	223	9	US-09-813-453A-67	Sequence 67, Appl
36	84	6.4	225	9	US-09-738-626-4158	Sequence 4158, Ap
37	84	6.4	417	10	US-09-828-3113-31	Sequence 31, Appl
38	81.5	6.2	258	9	US-09-738-626-5791	Sequence 5791, Ap
39	81	6.1	610	10	US-09-815-242-11183	Sequence 11183, A
40	81	6.1	1019	1	US-08-834-705-18	Sequence 18, Appl
41	80.5	6.1	380	10	US-09-815-242-10609	Sequence 10609, A
42	80.5	6.1	1864	9	US-09-832-292-27	Sequence 27, Appl
43	80.5	6.1	1885	10	US-09-920-346-2	Sequence 2, Appl
44	80	6.1	732	9	US-10-008-353-6	Sequence 6, Appl
45	79.5	6.0	435	10	US-09-815-242-12025	Sequence 12025, A

ALIGNMENTS

RESULT 1  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match	100.0%;	Score 1321;	DB 9;	Length 258;
Best Local Similarity	100.0%;	Pred. No. 4.3e-126;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSKTEDEYGMKALLNHVGLQFSDIRGI	60	
Db	1	MIFVLDVGNNTVLGVYDGDDELKHHWRIETSKTEDEYGMKALLNHVGLQFSDIRGI	60	
Qy	61	IISSVVPPIFALERMCLKYFHIKPLIVGPGIKTGLDKYDNPREGADRVNAVAGIHL	120	
Db	61	IISSVVPPIFALERMCLKYFHIKPLIVGPGIKTGLDKYDNPREGADRVNAVAGIHL	120	
Qy	121	YGSPLIIVDFCTATTTCYINEHKQYMGGAIFPGIMSTEALFARAAKLPRIETARDDII	180	
Db	121	YGSPLIIVDFCTATTTCYINEHKQYMGGAIFPGIMSTEALFARAAKLPRIETARDDII	180	
Qy	181	GKNTVSAMQAGILYGVQVEGTIVSRMAKSKIPPVKVIATGGLAPIASESDIIDVDFD	240	

Db 181 GKNTVSAMQAGILYGVQVEGIVSRMKA<sup>K</sup>SKIPKVIATGGLPIASESDIIDWD<sup>F</sup> 240

Qy 241 LTLTGLKILLYEK<sup>N</sup>TEKKG 258  
|||||

Db 241 LTLTGLKILLYEK<sup>N</sup>TEKKG 258  
|||||

RESULT 2  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent NO. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: C82-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

RESULT 3  
US-09-813-453A-2  
US-09-813-453A-2, Application US/09813453A  
; Sequence 2, Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OCG-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

[illegible]

RESULT 4  
US-09-813-453A-47  
: Sequence 47, Application US/09813453A  
: Patent No. US20020168681A1  
: GENERAL INFORMATION:  
: APPLICANT: Yocum, R. Rogers  
: APPLICANT: Patterson, Thomas A.  
: TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
: TITLE OF INVENTION: ANTIBIOTICS  
: FILE REFERENCE: OGZ-001  
: CURRENT APPLICATION NUMBER: US/09/813,453A  
: CURRENT FILING DATE: 2001-03-20  
: PRIOR APPLICATION NUMBER: US 60/227,860  
: PRIOR FILING DATE: 2000-08-24  
: PRIOR APPLICATION NUMBER: 09/667,569  
: PRIOR FILING DATE: 2000-09-21  
: NUMBER OF SEQ ID NOS: 77  
: SOFTWARE: Patentin ver. 2.0  
: SEQ ID NO 47  
: LENGTH: 254  
: TYPE: PRT  
: ORGANISM: Bacillus halodurans  
US-09-813-453A-47

	Query Match	73.8%	Score 975;	DB 9;	Length 254;
	Best Local Similarity	70.8%;	pred. No. 5.6e-91;		
	Matches 179;	Conservative 36;	Mismatches 38;	Indels 0;	Gaps 0;
Qy	1	MIFVLDGNTNTVLGVYDGDDELAKHHRIETSRSKTDEYCGMMTKALLNHWGLQFSDIRGI	60		
Db	1	MILVIDGNTNTVLGVYDQDTLVHHRLATSRKQTEDEYATVRSFDFHAGLQFQDIDGI	60		
Qy	61	IISVVPPIFMALERCLKTHFIKPLTVGIGIKTGLDIDKYNPREVGADRVNNAVAGIHL	120		
Db	61	VISSVVPMMFSLQCKKVFHYTPMIIIGPGIKTGLNIKYDNPKEVGADRVNNAVAIEL	120		
Qy	121	YGSPLIIVDFTGATTTCYINEHKQYMGGAITAPGIMISTEALFARAAKLPRIETARPDII	180		
Db	121	YGYPAIIVDFTGATTTCYINEKQYAGGVJAPGIMISTEALYHRAKSLPRIETIAKPKQV	180		
Qy	181	GKNTVSAMQAGILYGVGVGEGVTSRMKAQSKIPPVKVIATGGIAPLIASDESIDIIVVDPF	240		



Db 181 GTNIDSQSGIFGYGVQVGVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240  
QY 241 LTLGLKLLYEKN 253  
Db 241 LTLGLQLIYKKN 253

## RESULT 5

US-09-813-453A-17  
; Sequence 17, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-17

Query Match 58.1%; Score 900; DB 9; Length 233;  
Best Local Similarity 78.2%; Pred. No. 2e-83;  
Matches 169; Conservative 27; Mismatches 20; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MLLVIDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
Db 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
Db 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
QY 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP 216  
Db 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP 216

## RESULT 6

US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255

; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 63.1%; Score 834; DB 9; Length 255;  
Best Local Similarity 62.7%; Pred. No. 1.1e-76;  
Matches 160; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 1 MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MLLVIDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
Db 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
Db 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
QY 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP 240  
Db 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP 240  
QY 241 LTLGLKLLYEKNTE 255  
Db 241 LTLGLKLLYEKNTE 255

## RESULT 7

US-09-813-453A-55  
; Sequence 55, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 61.9%; Score 817.5; DB 9; Length 256;  
Best Local Similarity 62.9%; Pred. No. 5.4e-75;  
Matches 161; Conservative 40; Mismatches 54; Indels 1; Gaps 1;

QY 1 MIFVLDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
Db 1 MLLVIDVGNNTVLGVYDGDGLKHHWRIETSRKTSKDEYGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
Db 61 IISVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREGADRVNVAAGIHL 120  
QY 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
Db 121 YGSPLIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARAALKPRIETARPDDII 180  
QY 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP - KVIATGGLAPLIASESDI IDVDP 239  
Db 181 GKNVTSAQAGILYGVQVGVISVRMKAKSKIPP - KVIATGGLAPLIASESDI IDVDP 239

OY 240 FLTTLGLKLLYEKNT 255  
 DB 241 FLTLEGLRIIYKNOE 256

## RESULT 8

US-09-813-453A-4  
 ; Sequence 4, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 265  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces coelicolor  
 US-09-813-453A-4

Query Match 50.2%; Score 663; DB 9; Length 265;  
 Best Local Similarity 48.9%; Pred. No. 2.7e-59; Indels 8; Gaps 5;  
 Matches 129; Conservative 58; Mismatches 69; Indels 8; Gaps 5;  
 OY 1 MIFVLDVGNNTVGLVYDGDDELKHHRIETSRKTEDEYGMIMKALLN-H--VGLQFSD- 56  
 DB 1 MLLTIDVGNTHVGLGDEGIVHWRISTDSRRTADELAVLGLGMHPLGDELGDG 60  
 OY 57 IRGIISSVPPIMFALERMCKLYF-HIKPLIVGPGIKTGLDIKYDNPKEVGADRIYNAV 115  
 DB 61 IDGIAICATVPSVLEHREVTYYGDPVAVLVEPVKTVGPILTDHPKEVGADRIYNAV 120  
 OY 116 AGIHLGSLIIVDFGTATTCYINHEKQYMGGAATAPGIMISTEALFAAAKLPRIEIAR 175  
 DB 121 RAVELYGPAIVVDGTATTFDAVSARGEYIGGVITAPGIEISVEALGVGAQLRIEVAR 180  
 OY 176 PDDIIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPP---KVIATGGLAPLIASESD 232  
 DB 181 PRSVIGKNTVEAMQSGIVYGFAGQYDGVVNRMAELADDPDDVTVIATGGLAPVLGESS 240  
 233 IIDVVDPELTTLGLKLLYEKNT 256  
 DB 241 VIDEHEPWLTLGLRLVYERNVR 264

## RESULT 9

US-09-813-453A-3  
 ; Sequence 3, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3

; LENGTH: 250  
 ; TYPE: PRT  
 ; ORGANISM: Clostridium acetobutylicum  
 US-09-813-453A-3

Query Match 47.7%; Score 629.5; DB 9; Length 250;  
 Best Local Similarity 53.8%; Pred. No. 6.4e-56;  
 Matches 126; Conservative 40; Mismatches 59; Indels 9; Gaps 2;  
 OY 1 MIFVLDVGNNTVGLVYDGDDELKHHRIETSRKTEDEYGMIMKALLNHVGLQFSDIRGI 60  
 DB 18 VILVLDVGNNTVGLIYNDTKLTAENRSLDVLRSADYGIQVYNLFQDQKLDPTLVEGV 77  
 OY 61 IISVVPPIMFALERMCKLYFHKPLIVGPGIKTGLDIKYDNPKEVGADRIYNAVAGIHL 120  
 DB 78 IISVVPPIMYSLHMRIRYFKINPLVVGPGIKTGINIKYDNPKEVGADRIYNAVAHEI 137  
 OY 121 YGSLIIVDFGTATTCYINHEKQYMGGAATAPGIMISTEALFAAAKLPRIEIARPDII 180  
 DB 138 YKRSIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSSEALFEKAALPRVELIKPAYI 197  
 OY 181 GKNTVSAMQAGILYGVQVEGIVSRMKAKSKIP-----PKVIATGGLAPLI 227  
 DB 198 CKNTISSIQSGIVRYLRQVKLYFE--KLKENLPDGRTRTSLVLAATGGLAKLI 249

## RESULT 10

US-09-813-453A-51  
 ; Sequence 51, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 51  
 ; LENGTH: 260  
 ; TYPE: PRT  
 ; ORGANISM: Caulobacter crescentus  
 US-09-813-453A-51

Query Match 43.5%; Score 575; DB 9; Length 260;  
 Best Local Similarity 44.1%; Pred. No. 2.3e-50;  
 Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;  
 OY 1 MIFVLDVGNNTVGLVYDGDDELKHHRIETSRKTEDEYGMIMKALLNHVGLQFSDIRGI 60  
 DB 1 MLLAEQGNNTMTFAIHGASWVAQWRSATSESTRADYVYVWLSQLLSMQGLGFRAIDAV 60  
 OY 61 IISVVPPIMFALERMCKLYFHKPLIVGPGIKTGLDIKYDNPKEVGADRIYNAVAGIHL 120  
 DB 61 IISVVPSQIFNLNLSRRIYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAMV 120  
 OY 121 YGSLIIVDFGTATTCYINHEKQYMGGAATAPGIMISTEALFAAAKLPRIEIARPDII 178  
 DB 121 YGSLVVDISGTATTFDVAADGAFEGGIIAPGINSQALHAAAKLPRIAIOPAGNR 180  
 OY 179 IIGKNTVSAMQAGILYGVQVEGIVSRMKAKSKIPKPKVIATGGLAPLIASESDIIVVD 238  
 DB 181 IVGTDVTSAMQSGVFWGISLIEGLVARIKARGEPMVTIATGCVASLFEATDSIDHFD 240  
 OY 239 PELTTLGLKLLYEKNT 254  
 DB 241 SDLTIRGLLEIYRNRT 256

## RESULT 11

US-09-813-453A-6  
; Sequence 6, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Rhodobacter capsulatus  
US-09-813-453A-6

Query Match 41.0%; Score 542; DB 9; Length 258;  
Best Local Similarity 43.7%; Pred. No. 5.le-47;  
Matches 111; Conservative 47; Mismatches 94; Indels 2; Gaps 2;  
QY 1 MIFVLDGNTNTVLGVYDGDDELKHHRIETSRKTEDEYGMIMKALLNHVGLQFSDIRGI 60  
DB 1 MLCIDCGNTNTVFSVDWGDTFDAATWRIATDHRITADEYFVWLTMLQKGLQ-GRISEA 59  
QY 61 IISVVPPIFMALERMCLKYFHKPLIVG-PGKRTGLDIDKYDNPREVGVADRVNAVAGIH 119  
DB 60 IISSTAPRVNENLAVLCNRVDFCRPVYVKGCELPVAPRVDPGTGTVGPDRLVTVAGYD 119  
QY 120 LYGSPLIIVDFGTATTCYCYINEHKQYMGGAIPGIMISTEALFARAALPRLPIETARPDDI 179  
DB 120 RHGGDLIWDFTGATTEDVWAPDCAIGGVAPGNLSLEALHMAAALPHVDVTKPGV 179  
QY 180 IGKNTVSAMQAGILYGVGVQGVGVSRMKASKTPPKVIATGGLAPLIASESDIDVDVP 239  
DB 180 IGTNTVACIOSGVYWGIVGLVEGIVROIRMERDRMPKVIATGGLASLFDLGLDKVED 239  
QY 240 FLTLTGLKLLYEKN 253  
DB 240 DLTHWGLRLIFDYN 253

## RESULT 12

US-09-813-453A-57  
; Sequence 57, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 57  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Dehalococcoides ethenogenes

## US-09-813-453A-57

Query Match 38.8%; Score 513; DB 9; Length 219;  
Best Local Similarity 45.9%; Pred. No. 3.6e-44;  
Matches 96; Conservative 45; Mismatches 68; Indels 0; Gaps 0;  
QY 2 IFVLDGNTNTVLGVYDGDDELKHHRIETSRKTEDEYGMIMKALLNHVGLQFSDIRGII 61  
DB 5 LVAVDIGNTSVNIIGIFEKELANWHLGSAVQARMADAYASLLGLQHAGIHPEELNRVI 64  
QY 62 ISSVVPPIFMALERMCLKYFHKPLIVGPGIKTGLDIDKYDNPREVGVADRVNAVAGIHLX 121  
DB 65 MCSVVPPIPTTFEEYFKSYKAAPLVVGAGIKSGVAKVRMDNPREVGADRVNNAARVLY 124  
QY 122 GSPLIIVDFGTATTCYCYINEHKQYMGGAIPGIMISTEALFARAALPRLPIETARPDDI 181  
DB 125 PGACIIVDMGTATTFDLSGGAYIGGATAPGATSAQAIAEKTSKLPKIEIIRPAKVIG 184  
QY 182 KNTVSAMQAGILYGVGVQGVGVSRMKAK 210  
DB 185 SNTVSAMQSGIYFGYIGLVEELVRRIOTE 213

## RESULT 13

US-09-813-453A-5  
; Sequence 5, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-813-453A-5

Query Match 36.3%; Score 479.5; DB 9; Length 272;  
Best Local Similarity 37.8%; Pred. No. 1.2e-40;  
Matches 101; Conservative 56; Mismatches 93; Indels 17; Gaps 5;  
QY 1 MIFVLDGNTNTVLGVYDGDDELKHH-----WRIETSRKTEDEYGMIMKALLNHVGLQF 54  
DB 1 MLLAIDVNTHTVVGLLSG--MKEAKVVQWRITSEVTDELATIDGL---IGEDS 55  
QY 55 SDIRGIISVVPPIFMALERMCLKYFHKPL-IVGPGIKTGLDIDKYDNPREVGVADRVN 113  
DB 56 ERLGTAAALSTVPSVLHEVRIMLDQWPSVPHVLTPEVGTGIPLLVDNPKFVGADRVN 115  
QY 114 AVAGIHLGSPLIIVDFGTATTCYCYINEHKQYMGGAIPGIMISTEALFARAALPRLPI 173  
DB 116 CLAAFYDRFRKAAIIVDFGSSICVDVVSARKEGLGAIAPGVQVSSDAAAARARVEL 175  
QY 174 ARPDIDIGKNTVSAMQAGILYGVGVQGVGVSRMKA-----KSKIPPKVIATGGLAPLIA 228  
DB 176 APRSVGKNTVCEMAGAVFGAGLVGLVGRIRREDVSGFSVDHDAIVATGHTAPLL 235  
QY 225 SESDIDVDVDFPLTLTGLKLLYEKNTE 255  
DB 236 PELHTVDHYDQHLTLQGLRLYFERNLE 262

## RESULT 14

US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 36.0%; Score 475.5; DB 9; Length 272;  
Best Local Similarity 37.5%; Pred. No. 3.1e-40;  
Matches 100; Conservative 57; Mismatches 93; Indels 17; Gaps 5;  
QY 1 MIFVLDVGNNTVVLGVYDGDDELKHH-----WRIETSRSKTEDEYGMKIKALLNHVGLQF 54  
Db 1 VLLAIDVNTHTVVGLLSG--MKEHAKVVOQWRITSEVTADELALTIDGL---IGEDS 55  
QY 55 SDIRGIILSSVVPPIINFALERMCLKYFIHKP-LIVGPGIKTGIDIKYDNPREGVADRIYN 113  
Db 56 ERLTCTAALSTVPSVLHEVRIMLDQYWPSPHVLIEPGVGTGIPLLVDNPKVGVADRIYN 115  
114 AVAGIHLGSPLIIVDFGTATTCYINHEKQYMGGAIAPIGMISTEALFARAACLPRIEI 173  
116 CLAAAYDRFRKAAIVYDFGSSICVDVVSAGKEFLGGAIAPIGVQVSSDAAAARSAAALRRVEL 175  
QY 174 ARPDDIIGKNTVSAMQAGILYGVGVGEIVSRMKA-----KSKPPKVIATGGLAPLIA 228  
Db 176 APRSVVGNKNTVECMQAGAVFGAGLVGLVGIREDVSGFSDVDHDAIVATGHTAPLLL 235  
QY 229 SESDIIDVVDPPFLTTLTGLKLYEKNT 255  
Db 236 PELHVDVHDQHLTLQGLRLVFERNLE 262

RESULT 15  
US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8  
Query Match 31.5%; Score 415.5; DB 9; Length 262;  
Best Local Similarity 35.2%; Pred. No. 3.6e-34;  
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;  
QY 2 IFVLDVGNNTVVLGVYDGD-ELKHHWRIETSRSKTEDEYGMKIKALLNHVGLQFSDIRGI 60  
Db 6 LLAVDIGNTTTLVLGLADASGALHTWRIETSRSKTEDEYGMKIKALLNHVGLQFSDIRGI 63  
QY 61 LISSVVPPI-----MEALERMCLKYFIHKP-LIVGPGIKTGIDIKYDNPREGVADRIYN 116  
Db 64 VLSSVAPPVGENYALAKR-----HFMIDAFVSAENLPDVTVELDTPGSGVADRIYN 116  
QY 117 GIHLGSP-----LIIVDFGTATTCYINHEKQYMGGAIAPIGMISTEALFARAACL 168  
Db 117 ---LFGAEKYLGGLDYAVVYDFGTSTNFDVVGRRRFLGGILATGAQVSADALFARAACL 173  
QY 169 PRIETARPDDIIGKNTVSAMQAGILYGVGVGEIVSRMKA-----KSKPPKVIATGGLAPLIA 228  
Db 174 PRITLOAPETAIGKNTVHALQSGLVFGYAEWVGLLRIRRAELPGEAVAVATGGFSTVQ 233  
QY 229 SESDIIDVVDPPFLTTLTGLKLYEKNT 255  
Db 234 GICQEIYYDEITLTLGLVLELWASRSE 260

Search completed: June 24, 2003, 22:29:02  
Job time : 15.3382 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.6069 Seconds  
(without alignments)  
2342.388 Million cell updates/sec

Title: DS-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNSTTTLAIFTGD.....VIDELAVLRGSDLLLRNMP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	322	24.7	254	16 Q9KGH5	Q9KGH5 bacillus ha
2	307	23.6	258	2 Q9F985	Q9F985 bacillus st
3	307	23.6	259	16 Q92F54	Q92F54 listeria in
4	307	23.6	259	16 Q8YAC5	Q8YAC5 listeria mo
5	302	23.2	259	16 Q8XHL5	Q8XHL5 clostridium
6	295	22.6	265	16 Q9XBH6	Q9XBH6 streptomyce
7	288	22.1	256	16 Q8RFE4	Q8RFE4 fusobacteri
8	280.5	21.5	274	16 Q9CD56	Q9CD56 mycobacteri
9	274.5	21.1	272	16 Q06282	Q06282 mycobacteri
10	274	21.0	255	16 Q8R7M2	Q8R7M2 thermomanaer
11	268	20.6	273	16 Q97EB4	Q97EB4 clostridium
12	261	20.0	261	16 Q9A6Z1	Q9A6Z1 caulobacter
13	253.5	19.5	262	16 Q9RX54	Q9RX54 deinococcus
14	228.5	17.5	246	16 Q9WZY5	Q9WZY5 thermotoga
15	225	17.3	229	16 Q67753	Q67753 aquifex aeo
16	218	16.7	276	16 Q8YQD7	Q8YQD7 anaerobact

17	214	16.4	212	2 Q32514	Q32514 desulfovibr
18	197.5	15.2	273	16 Q83446	Q83446 treponema p
19	166.5	12.8	262	16 Q51477	Q51477 borrelia bu
20	162.5	12.5	592	16 Q9JXF1	Q9JXF1 neisseria m
21	159.5	12.2	592	16 Q9JW17	Q9JW17 neisseria m
22	157.5	12.1	295	16 Q8Y2M4	Q8Y2M4 raietonia s
23	157	12.0	248	16 Q9HWC1	Q9HWC1 pseudomonas
24	146.5	11.2	257	16 Q740A5	Q740A5 synecocyst
25	144	11.1	242	16 Q9PC14	Q9PC14 xylella fas
26	108	8.3	209	16 Q9PIA9	Q9PIA9 campylobact
27	105.5	8.1	223	16 Q25533	Q25533 helicobacte
28	99.5	7.6	223	16 Q92KY6	Q92KY6 helicobacte
29	99	7.6	224	16 Q98Q93	Q98Q93 mycoplasma
30	94.5	7.3	616	16 Q8UAV6	Q8UAV6 agrobacteri
31	94.5	7.3	657	16 Q8Y350	Q8Y350 raietonia s
32	93.5	7.2	271	2 Q9AMU1	Q9AMU1 bradyrhizob
33	93.5	7.2	2621	10 Q9LMZ3	Q9LMZ3 arabidopsis
34	93	7.1	296	12 Q91GD9	Q91GD9 epiphyas po
35	93	7.1	347	16 Q9ZJL9	Q9ZJL9 helicobacte
36	91	7.0	317	17 Q9HM88	Q9HM88 halobacteri
37	91	7.0	347	2 Q54484	Q54484 helicobacte
38	89.5	6.9	466	16 Q9A4A0	Q9A4A0 caulobacter
39	89.5	6.9	2874	12 Q9QDE2	Q9QDE2 cryptonectr
40	88.5	6.8	305	5 Q61800	Q61800 caenorhabdi
41	88	6.8	433	16 Q9RZJ8	Q9RZJ8 deinococcus
42	87.5	6.7	249	16 Q9PGS2	Q9PGS2 xylella fas
43	87.5	6.7	974	2 Q53943	Q53943 streptomyce
44	87.5	6.7	974	16 Q86683	Q86683 streptomyce
45	87.5	6.7	2658	10 Q9SGE4	Q9SGE4 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9KGH5	PRELIMINARY;	PRT;	254 AA.
AC	Q9KGH5			
DT	01-OCT-2000	(TRENBLrel. 15, Created)		
DT	01-OCT-2000	(TRENBLrel. 15, Last sequence update)		
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)		
DE	Hypothetical protein BH0086.			
GN	BH0086.			
OS	Bacillus halodurans.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Bacillaceae; Bacillus.			
OX	NCBI_TaxID=86665;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-C-125 / JCM 9153;			
RX	MEDLINE=20512582; PubMed=11058132;			
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,			
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,			
RA	Horikoshi K.			
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus			
RT	halodurans and genomic sequence comparison with Bacillus subtilis."			
RL	Nucleic Acids Res. 28:4317-4331(2000).			
DR	EMBL; AP001507; BAB03805.1; -			
DR	InterPro; IPR004619; Baf.			
DR	Pfam; PF03309; Bvg_acc_factor; 1.			
DR	TIGRFAMs; TIGR00671; baf; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 254 AA; 27907 MW; OE3B557BA7DAC176 CRC64;			
Query Match	24.7%;	Score 322;	DB 16;	Length 254;
Best Local Similarity	28.9%;	Pred No. 5.5e-21;		
Matches	79;	Conservative	54;	Mismatches 104;
				Indels 36;
				Gaps 4;
Qy	1 MRLVVDIGNSTTTLAIF-----	-----TGDEEPSVESVPSALFADSSMTREVF	43	
Db	1 MILVIDGNTVTLVGYQDETILVHRLATSRQKTEDEY-----	-----AMTVRSLSF	47	
Qy	44 GNMARKEGPEQAIACISVVPSPATAVGSALLESLSFVPLTICKLRPFRLDYATPHFTG	103		



RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durand L., Dussurget O.,  
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
RT "Comparative genomics of *Listeria species*,"  
RL Science 294:849-852(2001).  
DR EMBL; AL591974; CAD00748.1; --  
DR MEROPS; M41.009; --  
DR ListiList; LMO00221; --  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
QO SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;  
Query Match 23.6%; Score 307; DB 16; Length 259;  
Best Local Similarity 30.4%; Pred. No. 1.3e-19;  
Matches 83; Conservative 49; Mismatches 107; Indels 34; Gaps 7;  
QY 1 MRLVVDIGNTSTTALIF-----TGDEEPSVESVPSALFADSSMTREVFNGMARK 49  
DB 1 MRLVVDIGNTCTGVIEKQKLLKHWRMTDRHRTSDEL-----GMTVLNFFSYANLT 53  
QY 50 HGEPOAICSVPSATAVGSALLSFVPLTICCKLRPFRLDYATPHFGADRL-- 107  
DB 54 PSDIQGLIISVVPIMHAMETCMVYFNIRPLIVGPGIKTGLNKLKVDNPREIGSDRIYN 113  
QY 108 ALCAWSRHLFSEK----PVIADIGTAIFDVLDTVGNVYRGGLIMPIDMAGALHSRTAQ 164  
DB 114 AVAA-----SSEYGPVIVDFGTATTCYIDSGVYQGAIAFGIMISTEALYNRAK 167  
QY 165 LPQVRIDRPSLLGRSTTECKSGVFGVVKQIGGLVDLDAIRGDLVDFGSESVIVTGG 224  
DB 168 LPRVDIAESSQIGKSTVSSMQAGIFGVGQCEGII-----AEMKQSNASPV-VVATGG 222  
QY 225 NSRIIVPEIGPVSVIDELAVLRGSDLLLRMNP 257  
DB 223 LARMITKSSAVDILDPFLTKGLLELYRNKP 255  
RESULT 5  
Q8XHL5 PRELIMINARY; PRT; 259 AA.  
AC Q8XHL5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein CPE2468.  
GN CPE2468.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / TYPE A;  
RX PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic  
RT flesh-eater".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
DR EMBL; AP003194; BAB82174.1; --  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
DR PROSITE; PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28819 MW; OD5FEA3B7A145E10 CRC64;  
Query Match 23.2%; Score 302; DB 16; Length 259;  
Best Local Similarity 28.1%; Pred. No. 3.6e-19;  
Matches 74; Conservative 62; Mismatches 113; Indels 14; Gaps 6;  
QY 1 MRLVVDIGNTSTTALIFTGDEEPSVE-----SVPSALFAD--SSTMREVFNGMARKHGEPO 54  
DB 1 MILLIDVGNINVLGIH--DNEKYIASWRISTDSKTSDEYSIQVMOLFNOAKLNPEDEVE 58  
QY 55 AIAICSVPSATAVGSALLSFVPLTICCKLRPFRLDYATPHFGADRL--ALCAW 112  
DB 59 GLIISVVPNIMESLENMVKCKEPIVVGPGIKTGINKINDNPKVEGADRIVNAAF 118  
QY 113 SRHLSEKPVIAVDIGTAITFDVLDTVGNVYRGGLIMPIDMAGALHSRTAQLPQVRIDR 172  
DB 119 EXH---KKPMIIDFGTATTCATTEKGDYLGNCICPGIOISADALFERAAKLPRIELEK 175  
QY 173 PSLGRSTTECKSGVFGVVKQIGGLVDLDAIRGDLVDFGSESVIVTGGNSRIIVPE 232  
DB 176 PXSIVCKNTVTSMQAGIITGYIGKVEYIVRMRKEMM-DLCEKEPFLVATGGLAKLYSE 234  
QY 233 IGPVSVIDELAVLRGSDLLLRMN 255  
DB 235 TDVIDEVDKRLTLEGLKILYEN 257  
RESULT 6  
Q8XN6 PRELIMINARY; PRT; 265 AA.  
AC Q8XN6;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein SCO3380.  
GN SCO3380 OR SCE94.31C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RC Oliver K., Harris D.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RC Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger X., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces

RT coellicolor A3(2).  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL049628; CAB40880.1; .  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs: TIGR00671; baf; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 22.6%; Score 295; DB 16; Length 265;  
 Best Local Similarity 32.9%; Pred. No. 1.6e-18;  
 Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;

QY 1 MRLVVDIGNSTTTLAFTGDEEPSVE-----SVPSALFADSDST-----MREVFNGMAR 48  
 DB 1 MLTIIDVGNTHTVIGLFDG--EDIVEHWRISTDSRTADELAVLLOGLMGMHPLLG--- 55  
 QY 49 KHGEP-QAIAICSVPSAT-----AVGSALLES--LFSVPVLTICCKLRFPFR 93  
 DB 56 ELGDGIDGIAICATVPSVLHLEHREVTRRYGDPVAVLVEPGVKTGVPILT----- 105  
 / 94 LDYATHTTGADRLAICANSHLFSEKPIAVDIGTAITFDVLDVTGNYRGGLIMPIDM 153  
 DB 106 -DH--PREVGADRIINAVALYLG-GPAIVDFGTATPDVARSARGEYIGGVIAPGIEI 161  
 QY 154 MAGALSHRAQLPQVDRIDRPESLLGRSTTECIKSGVFWGVKQIGGLVDALRGDLVDFG 213  
 DB 162 SVEALGVKAQRKIEVARPSVIGNTVEAMQSGIVYGFAGQVDGVNRMARELADD-- 219  
 QY 214 ESTVEVITGNSRIIVPEIGPVSVIDE---LAVLRGSDLLLRNM 256  
 DB 220 PDDVTVIATGLAPMV---LGESSVIDEHEPWLTLMLRLVYERNV 262

RESULT 7  
 Q8RFE4 PRELIMINARY; PRT; 256 AA.  
 AC Q8RFE4  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Bvg accessory factor.  
 GN FN0761.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25586;  
 MEDLINE-21886394; PubMed-11889109;  
 KA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 A Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 RA Fonstein M., Kyrpides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586";  
 RL J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL: AE010586; AAL94957.1; .  
 KW Complete proteome.  
 SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 22.1%; Score 288; DB 16; Length 256;  
 Best Local Similarity 31.6%; Pred. No. 6.4e-18;  
 Matches 84; Conservative 43; Mismatches 117; Indels 22; Gaps 7;

QY 1 MRLVVDIGNSTTTLAFTGDEEPSVESPSALFADSDSTPRE-----VFGNMARKHGE--- 52  
 DB 1 MLIIGIDGNTH-----IVTGIYDNGELISTFRATNDKMTDEYFSYFNIT-KYNEISI 55  
 QY 53 --PQAIAICSVPSATAVGSALLESLSFVPLTICCKLRFPFLDYATPT-FGADRLAL 109  
 DB 56 KVDAILSSVVPNIILIIFFQFARKYFKVEATIVLEKLPFTPAKGINITFGADRIID 115

QY 110 CANSRHLFSEKPIAVDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSHRTAQLPOVR 169  
 DB 116 ITEAMQKYPDKNLVIFDFGTATTVLKK-GVYIGGILPGIDMSINALYGNATKLPVK 174  
 QY 170 IDRPESLLGRSTTECIKSGVFWGVKQIGGLVDALRGDLVDFGSESTVEVITGNSRII 229  
 DB 175 FTFPSSVLGDTMKQIOAIAIFFGVAGQIKHKIKKINEEL-----NEEFVLATGGLGKIL 229  
 QY 230 VPEIGPVSVIDEAVLRGSDLLLRNM 255  
 DB 230 SAEIDEIDEYDANLSLGLVLYKLN 255

RESULT 8  
 Q9CD56 PRELIMINARY; PRT; 274 AA.  
 ID Q9CD56  
 AC Q9CD56  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein ML0232.  
 GN ML0232.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TN;  
 MEDLINE-21128732; PubMed-11234002;  
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus";  
 RL Nature 409:1007-1011(2001).  
 DR EMBL: AL583917; CAC29740.1; .  
 DR Leproma; ML0232; .  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs: TIGR00671; baf; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 21.5%; Score 280.5; DB 16; Length 274;  
 Best Local Similarity 29.8%; Pred. No. 3.3e-17;  
 Matches 79; Conservative 53; Mismatches 120; Indels 13; Gaps 7;

QY 1 MRLVVDIGNSTTTLAFTGDEEPS--VES-----VPSALFADSDSTPREVFGNMARKHGEPO 54  
 DB 1 MLLAIDVRNTHTVVGLLSGSKHAKVYQWRIETSEVTADALAL--IIDGLIGDSDRL 58  
 QY 55 A-TAICSVPSATAVGSALLESLSF-SVPVLTICCKLRFPFLDYATPTFGADRLALCAW 112  
 DB 59 AGNAALSTVPSVLHVRIMLDQYWPSPVPHVLIIEPGVGTGPIPLVDNKEVGADRIVNCIA 118  
 QY 113 SRHLFSEKPIAVDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSHRTAQLPOVRIDR 172  
 DB 119 AFHKFQQAIV-YDFSSICVDVVSAKGEFLGGAIPGVQVSSDAAAARSALRRVELAR 177  
 QY 173 PSLGRSTTECIKSGVFWGVKQIGGLVDALRGDLVDFGE--STVEVITGNSRIIV 230  
 DB 178 PRSVGKNTVECMQAGVWFAGLVGLVGRMQDVEEFSGDLGNRVAVVATGHTAPLL 237  
 QY 231 PEIGPVSVIDEAVLRGSDLLLRNM 255  
 DB 238 PELHTVDHYDRHLTLHLGLRLVFPERN 262



```
RESULT 9
006282
ID O06282 PRELIMINARY; PRT; 272 AA.
AC O06282;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative,
DE Baf family).
GN RV3600C OR MTCY07H7B.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
X STRAIN=H37RV;
MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; 295557; CAB08944.1;
DR EMBL; AE007170; AAK48063.1;
DR TIGR; MT3706;
DR TuberculList; rv3600c; -.
DR InterPro; IPR004619; Baf.
JR Pfam; PF03309; Bvg_acc_factor; 1.
JR TIGRFams; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 29304 MW; 5D70E6E0F09AC8B CRC64;

Query Match 21.1%; Score 274.5; DB 16; Length 272;
Best Local Similarity 29.3%; Pred. No. 1.1e-16;
Matches 77; Conservative 52; Mismatches 125; Indels 9; Gaps 5;

QY 1 MRLVVDIGNTSTTLAFTGDEPS--VES----VPSALFADSTMTREVFNGMKGEPQ 54
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MLLAIDVNTVTVGLLSGMEKHAQVVOQWRIRTESEVTADELAL-TIDGLIGDSERLT 59
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 AIAICSVPSATAVGSALLESLF-SVPVLTICCKLRFFRDYATPHFTFGADRLALCAWS 113
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 GTAAISTVPSVLHVEVRIMLDQYWPSPVHLPEGVGTGIPLLVDNPKFEGADRVNCLAA 119
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 RHLFSEKPVIAVDIGTAITFDVLTGVNRYGRLMPGIDMMAGALHSRTAQLPQVRIDRP 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 YDRF-RKAAIVVDFGSSICVDVSAKGEFLGATAPGVQVSDAAARSALRRVELARP 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 ESSLGRSTTECIKSGFVGVKQIGGLVDVDAIRGDLVRDFGSESTVEIVTGGNSRIIVPEI 233
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 RSVVGKNTVECMQAGAVFGFAGLVGLVGRIRREDVSGFSVDHDAIVATGHTAPLLPEL 238
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 GVSVIDELAVLRGSDLLLRNMN 256
```

```
DB 239 HTVDHYDQHLFTLQGLRIVERNL 261
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
DE TTE2381.
GN Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AB013180; AAM25520.1;
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CAG6D CRC64;

Query Match 21.0%; Score 274; DB 16; Length 255;
Best Local Similarity 27.8%; Pred. No. 1.2e-16;
Matches 73; Conservative 53; Mismatches 119; Indels 18; Gaps 7;

QY 1 MRLVVDIGNTSTTLAFTGDEPSVPSALFADSTMTREVFNGMKG--HG----E 52
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MLLAIDVNTVTVGLLSGMEKHAQVVOQWRIRTESEVTADELAL-TIDGLIGDSERLT 56
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 POAIAICSVPSATAVGSALLESLF-SVPVLTICCKLRFFRDYATPHFTFGADRLALCAW 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 57 IDDVLISSVVPPLMTLQVMSLKRYFTKPIVVGPIKGINIKYDNPKEVGADRVNVA 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 SRHLFSEKPVIAVDIGTAITFDVLTGVNRYGRLMPGIDMMAGALHSRTAQLPQVRIDR 172
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 AYELYG-GPVIIVDFGTATTFCAISEKGEYLGIIAPGLMISADALFQRTAKLPKIDLT 175
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 PESLLGRSTTECIKSGFVGVKQIGGLVDVDAIRGDLVRDFGSESTVEIVTGGNSRIIVPE 232
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 176 PFTVINRNVTVASMQSGIIYGHV----GMVDYIVTRMKGEFAPSAY-VVATGGFANMIAEE 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 IGPVSVVIDELAVLRGSDLLLRNMN 255
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 SKTIDTVNEMLTLEGLRIIYERN 253
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 11
Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
DE CAC3200.
GN Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
```



QY 51 G-----EQATAICSVPSATAVGSALESLESFVPLVICCKLRFPRFLDYATPHTF 102  
Db 50 GLFTLAGAPIPRAAVLSSVAPVGENYALALKRHFMDAFVAENLPDVTVELDTGCV 109  
QY 103 GADRLA-LCAWSRHLSFKPVIADIGTAITFDVLTGVTGGLIMPIDMGALHSR 161  
Db 110 GADRLNFGAEKYLGLDYAVVDFTGTFVGRGRLGGILATGAQVSADALFAR 169  
QY 162 TAQIPQVDRIDRPSLLGRSTTECKSGVFWGVQVIGGLVDAIRGDLVRDGESEVIV 221  
Db 170 AAKLPRIPLQAPETAIGKNTVHALQSLGVFVGAEMVDGLLRIRRAELP---GEAV--AVA 224  
QY 222 TGGNSRIIVPEIGPVSVIDELAVLRG 247  
Db 225 TGGFSRTVOGICQEIYDEYDTLRLG 250

RESULT 14  
Q9WZY5  
Q9WZY5 PRELIMINARY; PRT: 246 AA.  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein TW0883.  
GN TW0883.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
Stewart A.M., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima";  
RL Nature 399:323-329(1999).  
DR EMBL: AE001754; AAD35964.1; -  
DR TIGR: TW0883; -  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs: TIGR00671; baf; 1.  
W Hypothetical protein; Complete proteome.  
Q SEQUENCE 246 AA; 27154 MW; 9E0309AD462CE266 CRC64;  
Query Match 17.5%; Score 228.5; DB 16; Length 246;  
Best Local Similarity 29.9%; Pred. No. 1.4e-12;  
Matches 79; Conservative 46; Mismatches 96; Indels 43; Gaps 12;  
QY 1 MRLVVDIGNTSTTIAIFGDEE-----PSVESVPSALFADSSMTREVFQGNMARKHGE 52  
Db 1 MYLLVDGNTSVFSI-TEDGKTRRLSTGVTQTELEF---SHLHPLLGDMAR---E 53  
QY 53 POATAICSVPSATAVGSALESLESFVPLVICCKLRFPRFLDYATPHTFGADRL 107  
Db 54 IKGIGVASVFTQNTVIERFSQKYPHIFWVKAANGCVK-----WNVKNPSEVGADRV 107  
QY 108 A-LCAWSRHLSFKPVIADIGTAITFDVLTGVTGGLIMPIDMGALHSRTAQLP. 166  
Db 108 ANVAVFKEY--GKNGIILDMGTATTVDLV-VNGSYEGAILPGFFMNVHSLFRGTAKLP 164  
QY 167 QVRIDRPSLLGRSTTECKSGVFWGVQVIGGLVDAIR---GDLVRDGESEVIVTG 223  
Db 165 LVKVPADVFWGKDPTEENIRLGVNNGSVYALEGIIRIKVYGLD-----PVVLTG 215  
QY 224 GNSRIIVPEIGPVSVIDELAVLRG 247  
Db 216 GQSK-IVKDMIKHEIFDEDLTKG 238

RESULT 15  
O67753  
ID O67753 PRELIMINARY; PRT: 229 AA.  
AC O67753;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein AQ\_1924.  
GN AQ\_1924.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
aeolicus";  
RL Nature 392:353-358(1998).  
DR EMBL: AE000763; AAC07720.1; -  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs: TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
Q SEQUENCE 229 AA; 26068 MW; 64C347A285FABE9A CRC64;  
Query Match 17.3%; Score 225; DB 16; Length 229;  
Best Local Similarity 27.7%; Pred. No. 2.6e-12;  
Matches 71; Conservative 39; Mismatches 104; Indels 42; Gaps 8;  
QY 3 LVVDIGNTSTTIAIFGDEEPSVESVPSALFADSSMTREVFQGNMARKHGEPOAIAICSV 62  
Db 4 LTVVDGNSVDIALWEGKKVKDFLKSHEEF-----LKEEF----- 39  
QY 63 PSATAVGSALLESLSF-----VPVLITICCKLRFPRFLDYATPHTFGADRLCAWSRHL 116  
Db 40 PKLKALGISVKQS-FSEKVRGKIPKIKFLKKNFPIQVDYKTPETLGTDRVAL-AYSARK 97  
QY 117 FSEKVPVIAVDICTATFDVLTGVTGVTGGLIMPIDMGALHSRTAQLPQVRIDRPSL 176  
Db 98 FYGNVWVISAGTALVIDLV-LEGKFGGFTILGLGKLKILSDLAEGIEPFFPEVEIF 156  
QY 177 LGRSTTECKSGVFWGVQVIGGLVDAIRGDLVRDGESEVIVITGNSRIIVPEIGPV 236  
Db 157 LGRSTRECVLGGAYRESTEFIKSTLKLWRVFRKPF-----KVITGGEKY-----PSKF 207  
QY 237 SVIDEAVLRGSDLL 252  
Db 208 GIYDPLLVHRGMNLL 223

Search completed: June 24, 2003, 22:16:16  
Job time : 24.6069 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 10.4272 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNSTTLAIFGTD.....VIDELAVLRSGLLRNMP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	7.5	782	4	US-09-543-084A-32
2	98	7.5	782	4	US-09-543-084A-33
3	98	7.5	782	4	US-09-543-084A-36
4	96	7.4	782	4	US-09-543-084A-34
5	96	7.4	782	4	US-09-543-084A-35
6	82	6.3	297	2	US-09-027-013-3
7	82	6.3	297	3	US-09-244-233-3
8	80.5	6.2	491	4	US-09-362-899-3
9	79.5	6.1	774	1	US-08-019-870-3
10	79	6.1	431	3	US-08-478-507-2
11	79	6.1	431	4	US-09-128-275A-2
12	79	6.1	431	4	US-09-553-427-2
13	79	6.1	991	4	US-09-877-730-12
14	79	6.1	1069	4	US-09-877-730-2
15	79	6.1	1072	4	US-09-877-730-18
16	79	6.1	1150	4	US-09-877-730-8
17	79	6.1	1693	3	US-08-478-507-7
18	79	6.1	1693	4	US-09-128-275A-7
19	79	6.1	1693	4	US-09-553-427-7
20	78.5	6.0	533	2	US-08-225-488-2
21	78.5	6.0	774	1	US-08-019-870-8
22	78	6.0	576	2	US-08-676-279-58
23	78	6.0	1891	2	US-08-804-227C-12
24	78	6.0	1891	2	US-08-804-198-6
25	77.5	5.9	430	4	US-09-134-001C-2981
26	77.5	5.9	773	1	US-08-019-870-1
27	77.5	5.9	773	1	US-08-019-870-6

28	77.5	5.9	774	1	US-07-747-901A-3	Sequence 3, Appl
29	77.5	5.9	774	1	US-07-935-312-3	Sequence 3, Appl
30	77.5	5.9	774	1	US-08-019-870-11	Sequence 11, Appl
31	77.5	5.9	774	1	US-08-633-760-44	Sequence 44, Appl
32	77.5	5.9	774	1	US-08-633-760-46	Sequence 46, Appl
33	77.5	5.9	774	1	US-08-633-760-50	Sequence 50, Appl
34	77.5	5.9	774	1	US-08-633-760-52	Sequence 52, Appl
35	76.5	5.9	609	4	US-08-969-683A-67	Sequence 67, Appl
36	76	5.8	3729	2	US-08-804-227C-4	Patent No. 5240706
37	75.5	5.8	600	6	5240706-1	Sequence 10, Appl
38	75.5	5.8	740	1	US-08-309-512-10	Sequence 10, Appl
39	75.5	5.8	740	5	PCT-US92-08756A-10	Sequence 21, Appl
40	75.5	5.8	774	1	US-08-314-309A-21	Sequence 21, Appl
41	75.5	5.8	1544	4	US-09-413-814-46	Sequence 9, Appl
42	74.5	5.7	652	2	US-08-426-125-9	Sequence 9, Appl
43	74.5	5.7	652	2	US-08-455-355-9	Sequence 22, Appl
44	74.5	5.7	712	4	US-09-877-730-22	Sequence 22, Appl
45	74.5	5.7	774	1	US-08-633-760-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1

US-09-543-084A-32  
; Sequence 32, Application US/09543084A  
; Patent No. 6361988

GENERAL INFORMATION:

APPLICANT: Frances H. Arnold

APPLICANT: Zhixin Shao

APPLICANT: Huimin Zhao

TITLE OF INVENTION: Recombination of Polynucleotide

SEQUENCE OF SEQUENCES: 36

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oppenheimer Wolff & Donnelly LLP

STREET: 2029 Century Park East, Suite 3800

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 98

SOFTWARE: Word 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/543,084A

FILING DATE: April 4, 2000

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/381,935

APPLICATION NUMBER: 60/041,666

FILING DATE: March 25, 1997

APPLICATION NUMBER: 60/045,211

FILING DATE: April 30, 1997

APPLICATION NUMBER: 60/046,256

FILING DATE: May 12, 1997

APPLICATION NUMBER: 08/905,359

FILING DATE: August 4, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Oldenkamp, David J.

REGISTRATION NUMBER: 29,421

REFERENCE/DOCKET NUMBER: 330187-89

TELECOMMUNICATION INFORMATION:

TELEPHONE: (310) 788-5000

TELEFAX: (310) 788-5100

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 782 amino acids

TYPE: amino acid

TOPOLOGY: linear

```
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 782 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-09-543-084A-33

Query Match          7.5%; Score 98; DB 4; Length 782;
Best Local Similarity 24.3%; Pred. No. 0.0072;
Matches 65; Conservative 32; Mismatches 75; Indels 96; Gaps 15;

QY 7 IGNTSTTLAFT--GDEE-----PSVESVPSALFADSSMTREVFNGNARKHGEPOA----- 55
DB 512 LGNERPRSLRNLGLDQIQORLAGTDLGPKGFTTARLQWVMEFN--RMHGAELARDDL 569
QY 56 IATCSVVPATAGSALLESFSPVLTICCKL--RPFRLDYATPHTFGADRLALCAWS 113
DB 570 VALCRQPATASNGAIVD-----LTAACIALSRFDERAD-----LDSRG 609
QY 114 RHLFSKPVIAVDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSRTAQLPQVRIDRP 173
DB 610 AHLTFEF-ALAGGIRFADTFEVTDPV-----RT-----P 637
QY 174 ESLLGRSTTECIKSGVFWGVVKGIGL-VDAIRGDLVRDF-----GES-TVEV 219
DB 638 RRL---NTTDPRTALADAVORLAGIPLDAKLGIHTDSRGERRIPIHGGRGEAGTFNV 694
QY 220 I-----VTGNSRIIVPEIGP 235
DB 695 IINPLVPGVGYPQVHGTSEFVMVAVELGP 722

RESULT 3
US-09-543-084A-36
; Sequence 36, Application US/09543084A
; Patent No. 6361988
; GENERAL INFORMATION:
;   APPLICANT: Frances H. Arnold
;   APPLICANT: Zhixin Shao
;   APPLICANT: Huimin Zhao
;   APPLICANT: Lorraine J. Giver
;   TITLE OF INVENTION: Recombination of Polynucleotide
;   TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
;   NUMBER OF SEQUENCES: 36
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
;     STREET: 2029 Century Park East, Suite 3800
;     CITY: Los Angeles
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 90067
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: Windows 98
;     SOFTWARE: Word 97
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/543,084A
;     FILING DATE: April 4, 2000
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 09/381,935
;     APPLICATION NUMBER: 60/041,666
;     FILING DATE: March 25, 1997
;     APPLICATION NUMBER: 60/045,211
;     FILING DATE: April 30, 1997
;     APPLICATION NUMBER: 60/046,256
;     FILING DATE: May 12, 1997
;     APPLICATION NUMBER: 08/905,359
;     FILING DATE: August 4, 1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Oldenkamp, David J.
;     REGISTRATION NUMBER: 29,421
;     REFERENCE/DOCKET NUMBER: 330187-89
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (310) 788-5000
;     TELEFAX: (310) 788-5100
```

REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 330187-89  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 788-5100  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-543-084A-36

Query Match 7.5%; Score 98; DB 4; Length 782;  
Best Local Similarity 24.3%; Pred. No. 0.0072;  
Matches 65; Conservative 32; Mismatches 75; Indels 96; Gaps 15;  
QY 7 IGNTSTTLAIFT--GDEE-----PSVESVPSALFADSSMTREVFNGNMARKHGEPOA---- 55  
DB 512 LGNERTPSRLRLGLDQIQORLAGTDGLPGKGTTLRLWQVMEFN--RMHGAELARDL 569  
QY 56 IACSVPSATVAGSALLESLFSPVLTICCKL--RFPFLDYATPHFTGADRILALCAWS 113  
DB 570 VALCRROPTATASNGAIVD-----LTAACIALSRFDERAD-----LDSRG 609  
QY 114 RHLSEKPVIAVDIGTATFDVLTGVNRYGGLIMPIDMAGALHSHRTAQLPQVRIDRP 173  
DB 610 AHLTFEF-ALAGGIRFADTFEVDPV-----RT-----P 637  
QY 174 ESLGRSTTECIKSGVFWGVVKQIGGL-VDAIRGDLVRDF-----GES-TVEV 219  
DB 638 RRL---NTDPRVRLTALADAVQRLAGIPLDAKLGDIHTDSRGERRIPHGGRGEAGTENV 694  
QY 220 I-----VTGNSRIIVPEIGP 235  
DB 695 ITNPLVPGVGYPPQVVGHTSFVMAVELGP 722

#### RESULT 4

US-09-543-084A-34  
Sequence 34, Application US/09543084A  
Patent No. 6361988

##### GENERAL INFORMATION:

APPLICANT: Frances H. Arnold

APPLICANT: Zhixin Shao

APPLICANT: Huimin Zhao

APPLICANT: Lorraine J. Giver

TITLE OF INVENTION: Recombination of Polynucleotide

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oppenheimer Wolff & Donnelly LLP

STREET: 2029 Century Park East, Suite 3800

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 98

SOFTWARE: Word 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/543.084A

FILING DATE: April 4, 2000

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/381,935

APPLICATION NUMBER: 60/041,666

FILING DATE: March 25, 1997

APPLICATION NUMBER: 60/045,211

FILING DATE: April 30, 1997

APPLICATION NUMBER: 60/046,256

FILING DATE: May 12, 1997  
APPLICATION NUMBER: 08/905,359  
FILING DATE: August 4, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Oldenkamp, David J.  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 330187-89  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 788-5100  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-543-084A-34

Query Match 7.4%; Score 96; DB 4; Length 782;  
Best Local Similarity 23.9%; Pred. No. 0.013;  
Matches 64; Conservative 32; Mismatches 76; Indels 96; Gaps 15;

QY 7 IGNTSTTLAIFT--GDEE-----PSVESVPSALFADSSMTREVFNGNMARKHGEPOA---- 55  
DB 512 LGNERTPSRLRLGLDQIQORLAGTDGLPGKGTTLRLWQVMEFN--RMHGAELVRDDL 569  
QY 56 IACSVPSATVAGSALLESLFSPVLTICCKL--RFPFLDYATPHFTGADRILALCAWS 113  
DB 570 VALCRROPTATASNGAIVD-----LTAACIALSRFDERAD-----LDSRG 609  
QY 114 RHLSEKPVIAVDIGTATFDVLTGVNRYGGLIMPIDMAGALHSHRTAQLPQVRIDRP 173  
DB 610 AHLTFEF-ALAGGIRFADTFEVDPV-----RT-----P 637  
QY 174 ESLGRSTTECIKSGVFWGVVKQIGGL-VDAIRGDLVRDF-----GES-TVEV 219  
DB 638 RRL---NTDPRVRLTALADAVQRLAGIPLDAKLGDIHTDSRGERRIPHGGRGEAGTENV 694  
QY 220 I-----VTGNSRIIVPEIGP 235  
DB 695 ITNPLVPGVGYPPQVVGHTSFVMAVELGP 722

#### RESULT 5

US-09-543-084A-35

Sequence 35, Application US/09543084A

Patent No. 6361988

GENERAL INFORMATION:

APPLICANT: Frances H. Arnold

APPLICANT: Zhixin Shao

APPLICANT: Huimin Zhao

APPLICANT: Lorraine J. Giver

TITLE OF INVENTION: Recombination of Polynucleotide

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oppenheimer Wolff & Donnelly LLP

STREET: 2029 Century Park East, Suite 3800

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 98

SOFTWARE: Word 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/543,084A

FILING DATE: April 4, 2000

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/381,935

APPLICATION NUMBER: 60/041,666  
FILING DATE: March 25, 1997  
APPLICATION NUMBER: 60/045,211  
FILING DATE: April 30, 1997  
APPLICATION NUMBER: 60/046,256  
FILING DATE: May 12, 1997  
APPLICATION NUMBER: 08/905,359  
FILING DATE: August 4, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Oldenkamp, David J.  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 330187-89  
TELEPHONE: (310) 788-5000  
TELEFAX: (310) 788-5100  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 782 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
-09-543-084A-35

Query Match 7.4%; Score 96; DB 4; Length 782;

Best Local Similarity 23.9%; Pred. No. 0.013;

Matches 64; Conservative 32; Mismatches 76; Indels 96; Gaps 15;

QY 7 IGWTTTIAFT--GDEE-----PSVESPSALFADSSSTMREVFNGMARKHG-----EPQA 55  
DB 512 LGNETPRSLRLGLDQIQRLAGTDLPGKGFTTARLQVNFNG--RMHGAELVRDDL 569  
QY 56 IALCSVPSATAVGSALESVPVLTICCKL--RPFRLDYATPHTFGADRLALCAWS 113  
DB 570 VALCRQPATASNAIVD-----LTAACIALSRFERAD-----LDSRG 609  
QY 114 RLHSEKPVIAVDIGTAIFDVLDTVGNVYRGGLIMPIDMAGALHSRTAQLPQVRIDRP 173  
DB 610 AHLTFEF-ALAGGIRFADTFETDPV-----RT-----P 637  
QY 174 ESLGRSTETCKSGVFWGVVQIGGL-VDAIRGDLVRDF-----GES-TVEV 219  
DB 638 RRL--NTDPRVTRALADAVQRLAGIPDLDAKLGDHTDSRGERRIPHGGREAGTFNV 694  
QY 220 I-----VTGGNSRIIVPEIGP 235  
DB 695 ITNPLVPGVGYPQVYVHGTSTFVMAVELGP 722

## RESULT 6

S-09-027-013-3

Sequence 3, Application US/09027013

Patent No. 5962302

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/027,013

FILING DATE: Herewith

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0462 US  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: GI 42131  
US-09-027-013-3

Query Match 6.3%; Score 82; DB 2; Length 297;

Best Local Similarity 22.2%; Pred. No. 0.14;

Matches 55; Conservative 34; Mismatches 91; Indels 68; Gaps 12;

QY 33 FADSSTMREVFNGMARKHGEPOAICSVVPSATAVGSALLESL--FSVPVLTICCKLRF 90  
DB 52 FVQSLSEREQVLEIVAEEGKGKIKLIARHVGCVTTAESQQLAASAKRYGFDVAVSVPFY 111  
QY 91 PFRDLATPH---TFGADRLALCAWSRHLFSEKPVIAVDI---GTAITFDVLDVTGNY 142  
DB 112 PFSFEHCDHYRAIIDSADGL-----PMVYNIPALSGVKLTLDQINTL--- 155  
QY 143 RGLIMPIDMAGALHSRTA---QLPQVRIDRPESLLGRSTTECIKSGVFWGVKQIG- 198  
DB 156 ---VTLPGV---GALKQTSGLDXQMQIRREHPDLYNGYDIFASGLLAGADGGIGS 208  
QY 199 -----GLVDAIR-GDLVRDFGSESTVEIVTGNRSRIIVPEIGPVSVIDEL---AV 244  
DB 209 TYNIMGWRYQGVKALKEGDI-----QTAQKLQTECN-----KVIDLLIKTV 251  
QY 245 LRGSDDL 252  
DB 252 FRGLKTVL 259

## RESULT 7

US-09-244-233-3

Sequence 3, Application US/09244233

Patent No. 6030824

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/244,233

FILING DATE:

CLASSIFICATION:



## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/027,013  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0462 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: GI 42131  
JS-09-244-233-3

Query Match 6.3%; Score 82; DB 3; Length 297;

Best Local Similarity 22.2%; Pred. No. 0.14;

Matches 55; Conservative 34; Mismatches 91; Indels 68; Gaps 12;

QY 33 FADSTREVEFGNMARKHGEPOAIAICSVVPSATAVGSALLESL--FSPVPLTICCKLRF 90  
DB 52 FVQSLSEOEVLVEIVAEKGGKILIAHVGCVTTAESQOLAASAKRYGFDASVATPFY 111  
QY 91 PFRIDYAPPH-----TFGADRLALCAWSRHLFSEKPVIAVDI---GTAITFDVLTGVNY 142  
DB 112 PFSPEHCDHYRAIIDSADGL-----PMVVYNIPALSKVLTLDQINTL--- 155  
QY 143 RGLIMPIDMAGALHSRTA---QLPQVRIDRPESLGRSTTECIKSGVFWGVKQIG- 198  
DB 156 ---VTLPGV---GALKOTSGDLYOMEQIRREHPDLVLYNGYDEIFASGLLAGADGGIGS 208  
QY 199 -----GLVDAIR--GDIVRDGCESTVEVIVTGGNSRIIVPEIGPVSVVIDEL---AV 244  
DB 209 TYNINGWYQIVKALKEGDI-----QTAQLQTECN-----KVIDLLIKTV 251  
QY 245 LRGSLL 252  
DB 252 FRGLKTVL 259

## RESULT 8

S-09-362-899-3  
Sequence 3, Application US/09362899  
Patent No. 6361986

## GENERAL INFORMATION:

APPLICANT: Degussa-Hls AG  
APPLICANT: Forschungszentrum-Jlich GmbH  
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND  
CURRENT APPLICATION NUMBER: US/09/362,899  
CURRENT FILING DATE: 1999-07-29  
EARLIER APPLICATION NUMBER: DE 19924365.4  
EARLIER FILING DATE: 1999-05-27  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3

LENGTH: 491

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-362-899-3

Query Match 6.2%; Score 80.5; DB 4; Length 491;

Best Local Similarity 21.5%; Pred. No. 0.48;

Matches 64; Conservative 47; Mismatches 111; Indels 75; Gaps 16;

QY 15 AIFTGDEEPSVESVPSA-LFADSTMTREVFGNMARKHGEPOAIAICSVVPSATAVGSALL 73  
DB 62 SVITG--EGTVGIPVAVLTSDSFLLGSLGTVA-----SVRINKAIHRAHTEKLP 112

QY 74 ESLES-----VPLTICCKLR-----FPERLDYATPHTFGADRLALCAW--S 113  
DB 113 VSPASGGARQEDNRAFVMMVSTAAVQRHREHLFP-LVILRNPTMG--ANASWGSS 168  
QY 114 RHLFSEKP-----VIAVDIGTAITFDVLTGVNRYRGLIMPIDMAGALHSRTA 163  
DB 169 GHLTFAEPGAQIGFLGPRVVELTTHALPDGVQQAENLVKTGV-----IDGIVSPLQLRAA 224  
QY 164 QLPQVRIDRPESLGR--STTECIKSGVFWGVKQ-----TGGLVDATRGDLVRDFGE 214  
DB 225 VAKTLKVIQVVEATDRFSTPTTGVALPVMFAIARSRDPPQPGIGETIMETLGADVVKLSGA 284  
QY 215 ST-----VEVITGGNSRIIVPE-----IGPVSVVIDELA-VLRGSDLLLRNMP 257  
DB 285 RAGALSPAVRVALARIGGRPVVLIGQDRRFTLGP-----QELRFARRGISLARELNLP 337

## RESULT 9

US-08-019-870-3

Sequence 3, Application US/08019870

Patent No. 5336613

## GENERAL INFORMATION:

APPLICANT: NIWA, MINO  
APPLICANT: YOSHIMASA, SAITO  
APPLICANT: SASAKI, HITOSHI  
APPLICANT: ISHII, YOSHINORI  
TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/019,870  
FILING DATE: 19930219  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Oblon, NO. 5336613man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-791-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-019-870-3

Query Match 6.1%; Score 79.5; DB 1; Length 774;

Best Local Similarity 23.4%; Pred. No. 1.3;

Matches 62; Conservative 29; Mismatches 101; Indels 73; Gaps 13;

QY 1 MRLVVDIGNTSTTTLAIFTGDEEPSVPSALSALFADSTMTREVFGNMARKHGEPOAIAI-- 58  
DB 187 LKLYDDGGRD-LLCIPPGAERDLE-----ADLATRPADVALLKMGGSDADG 238  
QY 59 ----CSVPSATAVGSALL-----ESLFSVP-----VLTICKLRFPF-- 92  
DB 239 GSNNAVAPGRTATGRPILAGDPRHVFIPGMVAQHLHACDRFDMIGLTVPGVPGFPFHA 298

QY 93 ---RLDYATPHTF-----GADRLAL-----CAWSRHLESEKPVIAVDIGTA 130  
Db 299 HNGKVAISYTHAFMDLHDLYLEOFAGEGRTAREGNDPEPVASRDR-----IAVRGGAD 352  
QY 131 ITFDVLDITVGNTRGGIMPGIDMMAGALHSRTAQLPQ--VRIDRPESLLGRSTTECKSG 188  
Db 353 REEDIVET-----RHGPVIAGDPDGAALILRSVQFAETDLSDFCLTRMPGASTVAQLYDA 408  
QY 189 V-FWGVVKQ--TGGILVDAIRGDLVR 210  
Db 409 TRGWGLDHLNVAGDVAGSIGHLVR 433

## RESULT 10

US-08-478-507-2  
; Sequence 2, Application US/08478507  
; Patent No. 6120988  
; GENERAL INFORMATION:  
; APPLICANT: Reyes, Gregory R  
; APPLICANT: Yarbough, Patrice O  
; APPLICANT: Bradley, Daniel W  
; APPLICANT: Krawczynski, Krzysztof Z  
; APPLICANT: Tam, Albert  
; APPLICANT: Fry, Kirk E  
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted  
; TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,507  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/279,823  
; FILING DATE: 25-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/681,078  
; FILING DATE: 05-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/505,888  
; FILING DATE: 05-APR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/420,921  
; FILING DATE: 13-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/367,486  
; FILING DATE: 16-JUN-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/336,672  
; FILING DATE: 11-APR-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/208,997  
; FILING DATE: 17-JUN-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0183.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0880  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-478-507-2

Query Match 6.18; Score 79; DB 3; Length 431;  
Best Local Similarity 24.78; Pred. No. 0.6;  
Matches 48; Conservative 31; Mismatches 69; Indels 46; Gaps 12;

QY 48 RKH-GEPOAIAICSVPSATAVGSALLESLSFVPLATICCKLRFPFLDYATPHTFGADR 106  
Db 263 KKHSGEP-----GTLNNTVNMVAITHC-----YDFR-DFQVAAFKGDDS 302  
QY 107 LALCAWSRHLESEKPVIAVDI-GTAITFDV-LDTVGNRYRGLIMPIDMM-----AGAL 158  
Db 303 IVLCSSEYR-----QSPGAALVIAGCGKLVDFRIGLYAGVVVAPGLGALPDVVRFAGRL 358  
QY 159 HSRT-----AQLPQVRIDRPESL--LGRSTTECK-----SGVFWGVVKQIGGLVDAIRG 206  
Db 359 TEKNWGPGERAEQLRLAVSDFLRKLTNVAQMCVDVVSRYVGSPLVHNLIGMLQAV-A 417  
QY 207 DLVRDFGESTVEVI 220  
Db 418 DGRAHFTESVKPVL 431

## RESULT 11

US-09-128-275A-2  
; Sequence 2, Application US/09128275A  
; Patent No. 6229005  
; GENERAL INFORMATION:  
; APPLICANT: Reyes, Gregory R  
; APPLICANT: Yarbough, Patrice O  
; APPLICANT: Bradley, Daniel W  
; APPLICANT: Krawczynski, Krzysztof Z  
; APPLICANT: Tam, Albert  
; APPLICANT: Fry, Kirk E  
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted  
; TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/128,275A  
; FILING DATE: 03-AUG-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/279,823  
; FILING DATE: 25-JUL-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/681,078  
; FILING DATE: 05-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/505,888  
; FILING DATE: 05-APR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/420,921  
; FILING DATE: 13-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/367,486  
; FILING DATE: 16-JUN-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/336,672

FILING DATE: 25-JUL-1994  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/681,078  
 FILING DATE: 05-APR-1991  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/505,888  
 FILING DATE: 05-APR-1990  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/420,921  
 FILING DATE: 13-OCT-1989  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/367,486  
 FILING DATE: 16-JUN-1989  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/336,672  
 FILING DATE: 11-APR-1989  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 07/208,997  
 FILING DATE: 17-JUN-1988  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sholtz, Charles K.  
 REGISTRATION NUMBER: 38,615  
 REFERENCE/DOCKET NUMBER: 4600-0183-22  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 324-0880  
 TELEFAX: (650) 324-0960  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 431 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-553-427-2

```

Db      263  KKHSGE-----GTLNWTVMNAVITHC-----YDFR-DFQVAAFKGDD 302
QY      107  LALCAWSRHLESEKPVIAVDI-GTAITFDV-LDVTGVNRYGGLIMPGIDMM-----AGAL 158
Db      303  IVLCSYR-----QSPGAAVLIAGCGLKLKVDFRPIGLYAGVVVAPGLGALPDVWRFAGRL 358
QY      159  HSRT-----AQLPOVRIDRPESL--LGRSTTECIK-----SGVFMGVVYKQJGLGVDAIRG 206
Db      359  TEKNWGPGRAEQGLRLAVSDFLRKLTNVAQCMVDVVSRYVSGPGLVHNLIQMLQAV-A 417
QY      207  DLYVRDFGCESTVEVI 220
Db      418  DGRAHFTESVKPVL 431

RESULT 13
US-09-877-730-12
; Sequence 12, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides En
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.2826 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNTSTTIAIFG.....VIDELAVLRSGLLLRMNP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 41779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_AA.\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303	100.0	257	9	US-09-813-453A-53
2	322	24.7	254	9	US-09-813-453A-47
3	319	24.5	258	9	US-09-813-453A-2
4	316	24.3	262	9	US-09-813-453A-45
5	305	23.4	258	9	US-09-813-453A-49
6	301	23.1	255	9	US-09-813-453A-7
7	295	22.6	265	9	US-09-813-453A-4
8	274.5	21.1	272	9	US-09-813-453A-5
9	273.5	21.0	233	9	US-09-813-453A-17
10	271.5	20.8	272	9	US-09-712-363-276
11	265.5	20.4	219	9	US-09-813-453A-57
12	265	20.3	238	9	US-09-813-453A-6
13	261	20.0	260	9	US-09-813-453A-51
14	254.5	19.5	256	9	US-09-813-453A-55
15	253.5	19.5	262	9	US-09-813-453A-8
16	229.5	17.6	250	9	US-09-813-453A-3
17	228.5	17.5	246	9	US-09-813-453A-9
18	227	17.4	241	9	US-09-813-453A-63
19	225	17.3	229	9	US-09-813-453A-12

20	221	17.0	244	9	US-09-813-453A-41	Sequence 41, Appl
21	214	16.4	212	9	US-09-813-453A-59	Sequence 59, Appl
22	197.5	15.2	273	9	US-09-813-453A-10	Sequence 10, Appl
23	166.5	12.8	262	9	US-09-813-453A-11	Sequence 11, Appl
24	162.5	12.5	592	9	US-09-813-453A-43	Sequence 43, Appl
25	161.5	12.4	460	9	US-09-813-453A-39	Sequence 39, Appl
26	159.5	12.2	592	9	US-09-813-453A-22	Sequence 22, Appl
27	157	12.0	248	9	US-09-813-453A-20	Sequence 20, Appl
28	153.5	11.8	249	9	US-09-813-453A-70	Sequence 70, Appl
29	146.5	11.2	257	9	US-09-813-453A-13	Sequence 13, Appl
30	144	11.1	242	9	US-09-813-453A-65	Sequence 65, Appl
31	136.5	10.5	267	9	US-09-813-453A-15	Sequence 15, Appl
32	130.5	10.0	249	9	US-09-813-453A-61	Sequence 61, Appl
33	108	8.3	209	9	US-09-813-453A-21	Sequence 21, Appl
34	105.5	8.1	223	9	US-09-895-913A-74	Sequence 74, Appl
35	105.5	8.1	223	9	US-09-813-453A-14	Sequence 14, Appl
36	105.5	8.1	223	9	US-09-813-453A-67	Sequence 67, Appl
37	98	7.5	787	10	US-09-205-448-8	Sequence 8, Appl
38	84	6.4	574	9	US-09-764-868-774	Sequence 774, Appl
39	83.5	6.4	588	10	US-09-815-242-5140	Sequence 5140, Ap
40	82	6.3	831	10	US-09-815-242-10520	Sequence 10520, A
41	30.5	6.2	491	9	US-09-738-626-4429	Sequence 4429, Ap
42	30.5	6.2	491	12	US-10-024-370-3	Sequence 3, Appl
43	80	6.1	5877	9	US-10-142-515-11	Sequence 11, Appl
44	80	6.1	5935	9	US-10-243-243A-8	Sequence 8, Appl
45	79	6.1	991	9	US-10-231-353-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-53  
; Sequence 53, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Chlorobium tepidum  
US-09-813-453A-53

Query Match	100.0%	Score 1303;	DB 9;	Length 257;
Best Local Similarity	100.0%;	Pred. No. 1.7e-126;		
Matches 257;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRLVVDIGNTSTTIAIFTGDEEPSVESVPSALFADSDSTMREVFGNNARKHGEPOAIAICS	60	
DB	1	MRLVVDIGNTSTTIAIFTGDEEPSVESVPSALFADSDSTMREVFGNNARKHGEPOAIAICS	60	
QY	61	VFSATAVGSALESLESFVPVLTICCKLRFPRFDYATPHFTFGADRLALCAWSRHLFSEK	120	
DB	61	VFSATAVGSALESLESFVPVLTICCKLRFPRFDYATPHFTFGADRLALCAWSRHLFSEK	120	
QY	121	PVIAVDIGTAITFDVLTGVNRYRGLIMPIDMAGALHSRTAQLPQVRIDRPSLLGRS	180	
DB	121	PVIAVDIGTAITFDVLTGVNRYRGLIMPIDMAGALHSRTAQLPQVRIDRPSLLGRS	180	
QY	181	TTECIKSGVFWGVVKQIGGLVDATRGDLVRDFGSEVTVITGNSRIIVEIGPVSVID	240	
DB	181	TTECIKSGVFWGVVKQIGGLVDATRGDLVRDFGSEVTVITGNSRIIVEIGPVSVID	240	



```
Db 176 PSSVVGKNTVSAMSGILYGVQGVGIVKRMKEE-----AKQEPKVIATGGLAKLISEE 230
; QY 233 IGPVSVIDELAVLRGSDLLLRN 255
; Db 231 SNVIDVDPFLTLKGLMYLYERN 253

RESULT 5
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49
```

```
Query Match 23.4%; Score 305; DB 9; Length 258;
Best Local Similarity 29.3%; Pred. No. 1.9e-23;
Matches 76; Conservative 52; Mismatches 121; Indels 10; Gaps 4;

QY 1 MRLVVDIGNTTTIAFTGDE---EPSVESVPSALFADSSM-REVFGNMARKHGEPOAI 56
; Db 1 MIFLDVGNNTVLGVGDYDELKHHWRIETSRKSTDEYEGMMIKALLNHVGLQFDIRGI 60

QY 57 AICSVVPSATAVGSALLESLSVPLTICCKLRFFRLDYATPHFTFGADRLCALCWSRHL 116
; Db 61 IISVVPPIMFALERMCLKYHFKPLIVGPGIKTGDKIDKIDNPREVGADRVNAVAGIHL 120

QY 117 FSEKPVIAVDIGTATFDVLTGVNRYGGLIMPIDMMAGALHSRTAQLPQVRIDRPESL 176
; Db 121 YG-SPLIIVDGTATYCYINEHKQYMGGAIPGIMISTEALFARAALPRIEIRPDDI 179

QY 177 LGRSTTECKSGVFWGVQKIGGLVDIAIRGDLVRDFGSEVIVTGNNSRIIVPEIGPV 236
; Db 180 IGKNTVSAMQAGILYGVQGVGIVSRMKAK-----SKIPPKVIATGGLAPLIASESDII 234

QY 237 SVIDEAVLRGSDLLLRN 255
; Db 235 DVVDPFLTLKGLMYLYERN 253
```

```
RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
```

```
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 23.1%; Score 301; DB 9; Length 255;
Best Local Similarity 31.8%; Pred. No. 4.9e-23;
Matches 85; Conservative 48; Mismatches 108; Indels 26; Gaps 8;

QY 1 MRLVVDIGNTTTIAFTGDEEPSVESVPSALFADSSMTREVFGE---NWKHKG-----E 52
; Db 1 MLLVDVGNNTVLGVGDYDELKHHWRIETSRKSTDEYEGMMIKALLNHVGLQFDIRGI 56

QY 53 POAIAICSVVPSATAVGSALLESLSVPLTICCKLRFFRLDYATPHFTFGADRL--ALC 110
; Db 57 IRAVLISSVVPPLTGVLEKLSGLYFGMRPLVVGPGIKTGMPIDNPREVGADRIYNVA 116

QY 111 AMSRHLFSEKPVIAVDIGTATFDVLTGVNRYGGLIMPIDMMAGALHSRTAQLPQVRI 170
; Db 117 GYEKYRTS---LIIVDFGTATFDVYVNRKGEYCGGAIPGLVISTEALFORASKLPVRI 173

QY 171 DRPESLGRSTTECKSGVFWGVQKIGGLVDIAIRGDLVRDFGES--TVEVIVTGNNSRI 228
; Db 174 IRPSAIIARTVNSMQAGIYGVYV---GLVDEI---VTRMKAESKDAPRVATGGLASL 226

QY 229 IYPEIGPVSVIDEAVLRGSDLLLRN 255
; Db 227 IAPESKTIEAVEEYLTLEGLIRLYERN 253
```

```
RESULT 7
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-09-813-453A-4
```

```
Query Match 22.6%; Score 295; DB 9; Length 265;
Best Local Similarity 32.9%; Pred. No. 2.1e-22;
Matches 94; Conservative 42; Mismatches 96; Indels 54; Gaps 13;

QY 1 MRLVVDIGNTTTIAFTGDEEPSVE---SVPSALFADSSMT-----MREVFGNMAR 48
; Db 1 MLLTIDVGNTHTVLGLFDG--EDIVEHWRIETSRRTADELAVLQGLMGHPLGLD--- 55

QY 49 KNGEP-QATAICSVVPSAT-----AVGSALLES--LFSVPLVICCKLRPFPR 93
; Db 56 ELGDGIDGIAICATVPSVLHLEHRETVRRYGVDPVAVLVEPGVKTGVPILT----- 105

QY 94 LDYATPHFTFGADRLCALCWSRHLFSEKPVIAVDIGTATFDVLTGVNRYGGLIMPIDM 153
; Db 106 -DH--PREVGADRIINAAVAVELYG-GPAIVDFGTATTTDAVSARGEYIGGVATPGIEI 161
```

QY 154 MAGALHSTRQAOLPOVRIDRPSLLGRSTTECKSGVFWGVVQKIGLVDAIRGDIYRDFG 213  
DB 162 SVEALGVKAQLKIEVARPSVIGNTVEAMQSGVIGYFAGQVDGVVNNMARELADD-- 219  
QY 214 ESTVEVITGNSRIIVPEIGVPSVIDE---LAVLRGSDLLLRNM 256  
DB 220 PDDVTVIATGGLAPMV---LGESSVIDEHEPWLTMLGLRLVYERNV 262

## RESULT 8

US-09-813-453A-5  
; Sequence 5, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 5  
; TYPE: PRT  
; LENGTH: 272  
; ORGANISM: Mycobacterium tuberculosis  
US-09-813-453A-5

Query Match 21.1%; Score 274.5; DB 9; Length 272;  
Best Local Similarity 29.3%; Pred. No. 2.9e-20;  
Matches 77; Conservative 52; Mismatches 125; Indels 9; Gaps 5;  
QY 1 MRLVVDIGNTSTTIAIFGTGDEEPS--VES---VPSALFADSTMRVEFGNARKGEPO 54  
DB 1 MLLAIDVRNTHVTVGLLSGMEHAKVQWQWRIRTESEVTADELAL-TIDGLIGESERT 59  
QY 55 AIAICSVPSATAVGSALLESF-SVPLTICCKLRFPRLDYATPHTFGADRLALCAWS 113  
DB 60 GTAALSTVPSLHVRYMLDQWPSVPHLIEPGVATGTPLLVDNKEVGADRVNCLAA 119  
QY 114 RHLSEKPVIAVDIGTAITFDVLDVGNRYRGLIMPGLDMMAGALHSRTAQLPOVRIDRP 173  
DB 120 YDRF-RKAAIVVDFGSSICVDVVSAGFELGGAIPGVQVSSDAAAARSALRRVELARP 178  
? 174 ELLGRSTTECKSGVFWGVVQKIGLVDAIRGDLVRDPEGSTVEVITGNSRIIVPEI 233  
DB 179 RSVGKNTVECMQAGAVFGAGLVDPGLVGRIRREDVSGFSVDHDAIVATGHTAPLLPEL 238  
QY 234 GPVSVIDELAVLRGSDLLLRNM 256  
DB 239 HTVDHYDQHLQLGLRLVFERNL 261

## RESULT 9

US-09-813-453A-17  
; Sequence 17, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-17

Query Match 21.0%; Score 273.5; DB 9; Length 233;  
Best Local Similarity 28.5%; Pred. No. 3e-20;  
Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;  
QY 1 MRLVVDIGNTSTTIAIF-----TGDEEPSVESVPSALFADSTMRVFF 43  
DB 1 MLLVVDVGNNTVLGVYHDKLEVHWRIETSRHKTEDEFGMI-----LRSLF 47  
QY 44 GNMARKGEPOAICSVPSATAVGSALLESFSPVLITICCKLRFPRLDYATPHTFG 103  
DB 48 DHSGLMFEQIDGIIISVWPPIPALERMCTKYFHPQIVGPGMK7GLNFKYDNPKEVG 107  
QY 104 ADRLALCAWSRHLFSEKPVIAVDIGTAITFDVLDVGNRYRGLIMPGLDMMAGALHSRTA 163  
DB 108 ADRIVNAVAAILYGNPLIVDFGTATTCYIDENKQYMGGAATPGTITISTEALYSRAA 166  
QY 164 QLPQVRIDRPSLLGRSTTECKSGVFWGVVQKIGLVDAIRGDLVRD 211  
DB 167 KLPRIETRPDNIIGNKTVSAMQSGILFVGVGVEGIVKRMKWAQKD 214

## RESULT 10

US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 20.8%; Score 271.5; DB 9; Length 272;  
Best Local Similarity 29.1%; Pred. No. 5.9e-20;  
Matches 76; Conservative 52; Mismatches 124; Indels 9; Gaps 5;





QY 52 EQQAIAICSVPSATAVGSALLSLESFVPLTII--CCKLRFPRLDYATPHTGADRLAL 109  
Db 57 -DAVIISVWPQSIFNLNLSRRYENVEPLVIGENAKLGIDVRIE--KPSEAGADRLVN 113  
QY 110 CAWSRHLSEKPIAVDICTAITFDVLDVGNVYRGGLIMPIDMAGALHSRTAQLPOVR 169  
Db 114 AIGAMVY-PGPLVIDSTATTFDIVADGAFEGGIIAPGINLSWQALHEAAKLPRIA 172  
QY 170 IDRP--ESLGRSTTECIKSGVFWGVYKQIGGLVDAIRGDLVRDFGSESTVEIVTGNRSR 227  
Db 173 IQRPAGNRIVGTDTVSAMQSGVFWGYSLEGLIVARIKAER---GE-PMTVIATGVAS 227  
QY 228 IIVPEIGPVSVIDELAVLRGSDLLLRN 255  
Db 228 LFEGATDSIDHFDSDLTIRGLLEIYRRN 255

## RESULT 14

US-09-813-453A-55

Sequence 55, Application US/09813453A

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

FILE REFERENCE: ANTIBIOTICS

CURRENT APPLICATION NUMBER: US/09/813,453A

CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

PRIOR FILING DATE: 2000-09-21

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 55

LENGTH: 256

TYPE: PRT

ORGANISM: Clostridium difficile

US-09-813-453A-55

Query Match 19.5%; Score 254.5; DB 9; Length 256;  
Best Local Similarity 28.6%; Pred. No. 3.1e-18;  
Matches 80; Conservative 47; Mismatches 102; Indels 51; Gaps 10;

QY 1 MRLVVDIGNTSTTLAFTGD-----EESVSE--SVPSALF-ADSSMTREVFGN 45  
Db 1 MLVVDVGNVGNVILYKGDVLYNRYWRIKTDREKTSDEYGLISNLFDFDNVNSIDI--- 57  
Y 46 MARKHGEPOAIAICSVPSATAVGSALLSLESFVPLTICCK-----LRFPRLDYA 97  
Db 58 -----DDVIISVWPN-----VMHSENF-CIKYCKQPLVGPGLTGLNIKYD 101  
QY 98 TPHTFGADRL--ALCAWSRHLSEKPIAVDICTAITFDVLDVGNVYRGGLIMPIDMA 155  
Db 102 NPKQVGADRIVNAVAGIEKY---GAPSLVDFGTATTFCATSEKGEYLGTTIAPGIKISS 158  
QY 156 GALSRTAQLPOVRIDRPESLLGRSTTECIKSGVFWGVYKQIGGLVDAIRGDLVRDFGES 215  
Db 159 EALFOSAKLPVRELAKPGMTCKSTVSAWQSGIIYGV---GLVDKIISIMKRELNC 214  
QY 216 TVEIVTGNRSRIIVPEIGPVSVIDELAVLRGSDLLLRN 255  
Db 215 DVKVIATGGLAKLIASETKSIDYVDGFLTLEGLRIIYERN 254

## RESULT 15

US-09-813-453A-8

Sequence 8, Application US/09813453A

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

FILE REFERENCE: ANTIBIOTICS

CURRENT APPLICATION NUMBER: US/09/813,453A

CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

PRIOR FILING DATE: 2000-09-21

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 262

TYPE: PRT

ORGANISM: Deinococcus radiopugnans

US-09-813-453A-8

Query Match 19.5%; Score 253.5; DB 9; Length 262;  
Best Local Similarity 27.8%; Pred. No. 4.1e-18;  
Matches 74; Conservative 46; Mismatches 103; Indels 43; Gaps 7;

QY 3 LVVDIGNTSTTLAFTGDEEPSVESFALFADSS-----TMRVEFGN--MARKH 50  
Db 7 LAVDIGNTTVLGL-----ADASGALTHTWRTNRNREMLPDDLALQLH 49  
QY 51 G-----EPQAIAICSVVPSATAVGSALLSLESFVPLTICCKLRFPRLDYATPHTF 102  
Db 50 GLFTLAGAPIPRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSV 109  
QY 103 GADRLA-LCAWSRHLSEKPIAVDICTAITFDVLDVGNVYRGGLIMPIDMAGALHSR 161  
Db 110 GADRLCNLFCAEYKLGGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFAR 169  
QY 162 TQALPOVRIDRPESLLGRSTTECIKSGVFWGVYKQIGGLVDAIRGDLVRDFGSESTVEIV 221  
Db 170 AAKLPRILOAPETAIGKNTVHALQSGLVFGYAEWVDGLLRIRRAELP---GEAV--AVA 224  
QY 222 TGNRSRIIVPEIGPVSVIDELAVLRG 247  
Db 225 TGGFSRTVQICQCEIDYDETLRLG 250

Search completed: June 24, 2003, 22:29:04  
Job time : 15.2826 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 29.091 Seconds  
(Without alignments)  
1177.181 Million cell updates/sec

Title: US-09-813-453A-53  
Perfect score: 1303  
Sequence: 1 MRLVVDIGNSTFLAFTGCD.....VIDELAVLRGSDLLLRMNP 257

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1303	100.0	257	AAU91174	Pantothenate kinas
2	322	24.7	254	AAU91171	Pantothenate kinas
3	319	24.5	258	AAU01243	B. subtilis novel
4	319	24.5	258	AAU91149	Bacillus subtilis
5	316	24.3	262	AAU91170	Pantothenate kinas
6	307	23.6	259	ABBA7661	Listeria monocytog
7	305	23.4	258	AAU91172	Pantothenate kinas
8	301	23.1	255	AAU91154	Geobacter sulfurre
9	295	22.6	265	AAU91151	Streptomyces coeli
10	274.5	21.1	272	AAU91152	Mycobacterium tube

11	273.5	21.0	233	AAU91163	Pantothenate kinas
12	271.5	20.8	272	AAU91163	Mycobacterium tube
13	265.5	20.4	219	AAU91176	Pantothenate kinas
14	265	20.3	258	AAU91153	Rhodobacter capsul
15	261	20.0	260	AAU91173	Pantothenate kinas
16	254.5	19.5	256	AAU91175	Pantothenate kinas
17	253.5	19.5	262	AAU91155	Deinococcus radiop
18	229.5	17.6	250	AAU91150	Clostridium acetob
19	228.5	17.5	246	AAU91156	Thermotoga maritim
20	227	17.4	241	AAU91179	Pantothenate kinas
21	225	17.3	229	AAU91159	Aquifex aeolicus p
22	221	17.0	244	AAU91168	Pantothenate kinas
23	214	16.4	212	AAU91177	Pantothenate kinas
24	197.5	15.2	273	AAU91157	Treponema pallidum
25	169	13.0	389	AAU74909	Neisseria meningit
26	166.5	12.8	262	AAU91158	Borrelia burgdorfe
27	162.5	12.5	592	AAV38615	Neisseria meningit
28	162.5	12.5	592	AAV74912	Neisseria meningit
29	162.5	12.5	592	AAU91169	Pantothenate kinas
30	161.5	12.4	455	AAV38617	Neisseria gonorrhoe
31	161.5	12.4	455	AAV74908	Neisseria gonorrhoe
32	161.5	12.4	460	AAU91167	Pantothenate kinas
33	161.5	12.4	592	AAV38618	Neisseria gonorrhoe
34	151.5	12.4	592	AAV74911	Neisseria gonorrhoe
35	159.5	12.2	455	AAV74910	Neisseria meningit
36	159.5	12.2	592	AAV38616	Neisseria meningit
37	159.5	12.2	592	AAV74913	Neisseria meningit
38	159.5	12.2	592	AAU91166	Pantothenate kinas
39	157	12.0	248	AAU91164	Pantothenate kinas
40	153.5	11.8	249	AAU91182	Pantothenate kinas
41	146.5	11.2	257	AAU91180	Synechocystis pant
42	144	11.1	242	AAU91180	Pantothenate kinas
43	136.5	10.5	267	AAU91162	Bordetella pertussis
44	130.5	10.0	249	AAU91178	Pantothenate kinas
45	119	9.1	189	AAV38614	Neisseria meningit

ALIGNMENTS

RESULT 1  
AAU91174  
ID AAU91174 standard; Protein; 257 AA.  
XX AC AAU91174;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #12.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Chlorobium tepidum.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX N-PSDB; ABK54195.  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 10; Page 103-104; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 257 AA;

Query Match 100.0%; Score 1303; DB 23; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-143;  
 Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MRLVVDIGNSTTLAFTGDEEPSVPSALFADSTMTREVFGNMARKHGEPOAIAICS 60  
 |||||  
 1 MRLVVDIGNSTTLAFTGDEEPSVPSALFADSTMTREVFGNMARKHGEPOAIAICS 60  
 |||||  
 61 VVPSATAVGSALESLSFVPLTICCKLRFPFLDYATHTFGADRLALCAWSRHLESEK 120  
 |||||  
 61 VVPSATAVGSALESLSFVPLTICCKLRFPFLDYATHTFGADRLALCAWSRHLESEK 120  
 |||||  
 121 PVIADVGTATFDVLDVTGNYRGGLIMPIDMAGALHSRTAQLPQVRIDRPSLLGRS 180  
 |||||  
 121 PVIADVGTATFDVLDVTGNYRGGLIMPIDMAGALHSRTAQLPQVRIDRPSLLGRS 180  
 |||||  
 181 TTECKSGVFWGVKIGGLVDAIRGDLVRDFGSESTVEIVTGGNSRIIVPEIGPVSV 240  
 |||||  
 181 TTECKSGVFWGVKIGGLVDAIRGDLVRDFGSESTVEIVTGGNSRIIVPEIGPVSV 240  
 |||||  
 241 ELAVLRGSDLLLRMNP 257  
 |||||  
 241 ELAVLRGSDLLLRMNP 257  
 |||||

RESULT 2  
 AAU91171  
 ID AAU91171 standard; Protein; 254 AA.  
 AC  
 AAU91171;  
 XX  
 05-JUN-2002 (first entry)  
 DT

Pantothenate kinase (Coax) #9.  
 XX  
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX  
 OS Bacillus halodurans.  
 XX  
 PN WO200216601-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 24-AUG-2001; 2001WO-US26531.  
 XX  
 PR 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX  
 PI Yocum RR, Patterson TA;  
 XX  
 DR WPI; 2002-269358/31.  
 DR N-PSDB; ABK54192.  
 XX

PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 10; Page 100; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 254 AA;

Query Match 24.7%; Score 322; DB 23; Length 254;  
 Best Local Similarity 28.9%; Pred. No. 6.6e-29;  
 Matches 79; Conservative 54; Mismatches 104; Indels 36; Gaps 4;  
 1 MRLVVDIGNSTTLAIF-----TGDEEPSVPSALFADSTMTREV 43  
 |||||  
 1 MILVIDGNTVTLGVQDETLLVHWRLATSRQKTEDEY-----AMTVRSLF 47  
 |||||  
 44 GNMARKHGEPOAIAICSVFSAVGSALLESLSFVPLTICCKLRFPFLDYATHTFG 103  
 |||||  
 48 DHAGLQPDIDGIVISSVPPMPSLEQMKCKYFHTVPMIIGPGIKTGNIKYDNPKEVG 107  
 |||||  
 104 ADRLALCAWSRHLESEKRPVIAVDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSRTA 163  
 |||||  
 108 ADRIVNAVAIAELYG-YPAIVVDFGTATTYCLINEKKOYAGGVIAPGIMISTEALYHRAS 166  
 |||||  
 164 QLPQVRIDRPSLLGRSTTECKSGVFWGVKIGGLVDAIRGDLVRDFGSESTVEIVTGG 223  
 |||||  
 167 KLPRIETAKPKQVVGTNTIDMSQSGIFGYVSQDGVVVKRMAQ-----ASEPKVIATG 221  
 |||||  
 224 GNSRIIVPEIGPVSVIDEALVLRGSDLLLRNM 256  
 |||||  
 222 GLAKLTSEETIDVIDSFLTLKGLLIYKNV 254  
 |||||

RESULT 3  
 AAU01243  
 ID AAU01243 standard; Protein; 258 AA.  
 AC  
 AAU01243;  
 XX  
 18-JUL-2001 (first entry)  
 DT  
 XX  
 DE B. subtilis novel pantothenate kinase encoded by the gene coax.  
 XX  
 KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
 KW nutritional supplement; panto-compound; pantoate.  
 XX  
 OS Bacillus subtilis.  
 XX  
 PN WO200121772-A2.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 21-SEP-2000; 2000WO-US25993.  
 XX  
 PR 21-SEP-1999; 99US-0400494.  
 PR 07-JUN-2000; 2000US-0210072.  
 PR 28-JUL-2000; 2000US-0221836.  
 PR 24-AUG-2000; 2000US-0227860.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS.  
 XX  
 PI Yocum RR, Patterson TA, Hermann T, Pero JG;

XX WPI; 2001-218644/22.  
DR N-PSDB; AAS00984.  
XX  
PT New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
XX  
XX Example 14; Fig 23; 292pp; English.  
PS The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coaX), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;  
Query Match 24.5%; Score 319; DB 22; Length 258;  
Best Local Similarity 28.1%; Pred. No. 1.5e-28;  
Matches 76; Conservative 52; Mismatches 106; Indels 36; Gaps 4;  
QY 3 LVDIGNSTTLAIF-----TGDEEPSVESVPSALFADSTMTREVFGN 45  
DB 3 LVIDGNTNTVLGVYHDGKLEWHRIETSRHKTEDEFGMI-----LRSLEFDH 49  
QY 46 MARKHGEPOAIAICVSATAVGSALLESLFSPVLITCCKLPFFRLDYATPHTFGAD 105  
DB 50 SGLMFEQIDGIISSVPPIMFALERMCTKYFHEIPQIVGPMKTLNICYDNPKVEVGAD 109  
QY 106 RLALCAWSRHLFSEKPVIAVDIGTAITFDVLTGVNRYGGLIMPIDMAGALHSRTAQL 165  
DB 110 RIVNAVAIAHLYG-NPLIWDVFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKL 168  
QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVQKIGGLVDALRGDLVRDFGSESTVEIVTGGN 225  
DB 169 PRIETRPDNIIGKNTVSAMOSGILFGYGVQVEGIVKRMKWAQKD-----LKVATGGL 223  
QY 226 SRIIPEIGPVSVIDEVLAVLRGSDLLLRN 255  
DB 224 APLIANESDCIDIVDPFLTKGLELIYERN 253  
RESULT 4  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX  
AC AAU91149;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Bacillus subtilis pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PA Yocum RR, Patterson TA;  
PI  
XX WPI; 2002-269358/31.  
DR N-PSDB; ABK54168.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
XX Claim 10; Page 67-68; 128pp; English.  
PS The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 258 AA;  
Query Match 24.5%; Score 319; DB 23; Length 258;  
Best Local Similarity 28.1%; Pred. No. 1.5e-28;  
Matches 76; Conservative 52; Mismatches 106; Indels 36; Gaps 4;  
QY 3 LVDIGNSTTLAIF-----TGDEEPSVESVPSALFADSTMTREVFGN 45  
DB 3 LVIDGNTNTVLGVYHDGKLEWHRIETSRHKTEDEFGMI-----LRSLEFDH 49  
QY 46 MARKHGEPOAIAICVSATAVGSALLESLFSPVLITCCKLPFFRLDYATPHTFGAD 105  
DB 50 SGLMFEQIDGIISSVPPIMFALERMCTKYFHEIPQIVGPMKTLNICYDNPKVEVGAD 109  
QY 106 RLALCAWSRHLFSEKPVIAVDIGTAITFDVLTGVNRYGGLIMPIDMAGALHSRTAQL 165  
DB 110 RIVNAVAIAHLYG-NPLIWDVFGTATTCYIDENKQYMGGAIAPGITISTEALYSRAAKL 168  
QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVQKIGGLVDALRGDLVRDFGSESTVEIVTGGN 225  
DB 169 PRIETRPDNIIGKNTVSAMOSGILFGYGVQVEGIVKRMKWAQKD-----LKVATGGL 223  
QY 226 SRIIPEIGPVSVIDEVLAVLRGSDLLLRN 255  
DB 224 APLIANESDCIDIVDPFLTKGLELIYERN 253  
RESULT 5  
AAU91170  
ID AAU91170 standard; Protein; 262 AA.  
XX  
AC AAU91170;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Pantothenate kinase (Coax) #8.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Bacillus anthracis.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX

PR 24-AUG-2000; 2000US-227860P.  
 XX 20-MAR-2001; 2001US-0813453.  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 PI WPI; 2002-269358/31.  
 DR N-PSDB; ABK54191.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 8; Page 98-99; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 262 AA;  
 Query Match 24.3%; Score 316; DB 23; Length 262;  
 Best Local Similarity 28.9%; Pred. No. 3.5e-28;  
 Matches 76; Conservative 56; Mismatches 113; Indels 18; Gaps 5;  
 Qy 1 MRLVVDIGNTSTTIAFTGDEEPSVESPSALFADSTMRVFGNMARK-----HGE 52  
 Db 1 MIFVLDVGNNTAVLGVF---BEGELRQ-HWRMETDRHKTDEYGLMKLLEHGLSPED 56  
 Qy 53 PQATAICSVSATAVGSALLESFSPVLITICCKLRPFRLDYATPHTFGADRLALCAW 112  
 Db 57 VKGIVSSVPPIMPALERMCKYFKIRPVGVPGIKTGLNPKYENPREVGADRVNAVA 116  
 Qy 113 SRHLSEKPVIAVDIGTAITFDVLTGVNRYGGLIMPIDMAGALHSRTAQLPQVRIDR 172  
 Db 117 GIHLVG-SPLIIVDFGTATTTCYINEKHVGGVITPGIMISAEALYSRAKLPRITWK 175  
 Qy 173 PESLLGRSTTECIKSVFVGWVKQIGLVDAIRGDLVRDGEVTEVIVTGNSRIIYPE 232  
 Db 176 PSSVVGKNTVSAMQSGILYGVQGVGIVKRMKEB-----AKQEPKVIATGGLAKLISEE 230  
 233 IGPVSVIDELAVLRGSDLLLRN 255  
 231 SNVIDVDVDFTLKGLYLYERN 253  
 RESULT 6  
 ID ABB47661  
 XX ABB47661 standard; Protein; 259 AA.  
 AC ABB47661;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #365.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX  
 PD 18-OCT-2001.  
 XX

PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Deboux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Anend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 DR WPI; 2002-010914/01.  
 XX  
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides  
 XX  
 PS Claim 6; SEQ ID No 366; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 259 AA;  
 Query Match 23.6%; Score 307; DB 23; Length 259;  
 Best Local Similarity 30.4%; Pred. No. 3.9e-27;  
 Matches 83; Conservative 49; Mismatches 107; Indels 34; Gaps 7;  
 Qy 1 MRLVVDIGNTSTTIAFTGDEEPSVESPSALFADSTMRVFGNMARK 49  
 Db 1 MILVIDGNTNCTVGVYKQKLKHWRMTDRHTSDEL-----GMTVNFYSYANLT 53  
 Qy 50 HGEPOAIAICSVVPSATAVGSALLESFSPVLITICCKLRPFRLDYATPHTFGADRL-- 107  
 Db 54 PSDIIGIILSSVPPIMHAMETMCVRYFNIRLIVPGIKTGLNKLKNDPREIGSDRVN 113  
 Qy 108 ALCANSRHLFSEK---PVIAVDIGTAITFDVLTGVNRYGGLIMPIDMAGALHSRTAQ 164  
 Db 114 AVAA-----SSEYGTVPVIVDFGTATTTCYIDGSGYOGGAIAPGIMISTEALYNRAK 167  
 Qy 165 LPQVRIDRPESILLGRSTTECIKSVFVGWVKQIGLVDAIRGDLVRDGEVTEVIVTGG 224  
 Db 168 LPRVDIAESSQIIGKSTVSSMOAGIFYGVQCEGII-----AEMKKOSNASPV-VVATGG 222  
 Qy 225 NSRIIVPEIGPVSVIDEALAVLRGSDLLLRNMP 257  
 Db 223 LARMITERSSAVDILDLPFTLKGLELYRNKP 255  
 RESULT 7  
 ID AAU91172  
 XX AAU91172 standard; Protein; 258 AA.

AC AAU91172;  
 XX 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #10.  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Bacillus stearothermophilus.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 XX 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX N-PSDB; ABK54193.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 101-102; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 258 AA;  
 SQ  
 Query Match 23.4%; Score 305; DB 23; Length 258;  
 Best Local Similarity 29.3%; Pred. No. 6,6e-27;  
 Matches 76; Conservative 52; Mismatches 121; Indels 10; Gaps 4;  
 QY 1 MRLVVDIGNTTTLAIFTGDE---EPSVESVPSALFADSSVM-REVFGNMARKHGEPOAI 56  
 DB 1 MIFVLDVGNNTVLGVGDGELKHHWRIETSKRTEDEYGMIRALLNHVGLQFSDIRGI 60  
 QY 57 AICSVPSATAVGSALESFSPVLTICCKLRPFRLDYATPHTFGADRALCAWSRHL 116  
 DB 61 IISVSVPIPFALERMCKLYEHIKPLIVGPQIKTGLDIKYNPREVGADRVNAVAGIHL 120  
 QY 117 FSEKPIAVDIGTAITFDVLDVTGNYRGLLMPGIDMMAGALHSRTAQLPOVRIDRPESL 176  
 DB 121 YG-SPLIVDGTATTTCYINEHKQYMGATAPGIMISTEALFARAALPRIETARPDII 179  
 QY 177 LGRSTTECIKSGVGVVQKIGGLVDAIRGDLVRDFGSESTVITGNSRIIVPEIGPV 236  
 DB 180 IGKNTVSAMQAGILYGVQVEGIVSRMKAK-----SKIPKVIATGGLAPLASESDII 234  
 QY 237 SVIDEALVLRGSDLLLRN 255  
 DB 235 DVDPDFLTLGLKLLYEKN 253  
 RESULT 8  
 AAU91154

ID AAU91154 standard; Protein; 255 AA.  
 XX AAU91154;  
 AC 05-JUN-2002 (first entry)  
 XX Geobacter sulfurreducens pantothenate kinase Coax.  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Geobacter sulfurreducens.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 XX 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 72-73; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 255 AA;  
 SQ  
 Query Match 23.1%; Score 301; DB 23; Length 255;  
 Best Local Similarity 31.8%; Pred. No. 1.9e-26;  
 Matches 85; Conservative 48; Mismatches 108; Indels 26; Gaps 8;  
 QY 1 MRLVVDIGNTTTLAIFTGDEPSPVESVPSALFADSSVMREVFQ---NMARKHG---E 52  
 DB 1 MLLVVDVGNNTVLGIYDGERLVRDWRVST---DKARTDEYGLINELFRLAGLGLDQ 56  
 QY 53 POAIAICSVPSATAVGSALESFSPVLTICCKLRPFRLDYATPHTFGADRL---ALC 110  
 DB 57 IRATVITSSVPPPLTGLERLSLGYFGMRPLWVGPGIKTGMPIQYDNPREGVADRVNVA 116  
 QY 111 AWSRHLFSEKPIAVDIGTAITFDVLDVTGNYRGLLMPGIDMMAGALHSRTAQLPOVRI 170  
 DB 117 GYEKRTS---LIIVDFGTATTFDYVNRKGEYCGGAIFGLVISTEALFORAKSLPRVDI 173  
 QY 171 DRPESILGRSTTECIKSGVGVVQKIGGLVDAIRGDLVRDFQES--TVEVITVGTGNSRI 228  
 DB 174 IRPSAIIANTVNSMQAGIYYGV-----GLVDEI---VTRMKAESKADPRVIATGGLASL 226  
 QY 229 IVEIGPVSVIDEALVLRGSDLLLRN 255  
 DB 227 TAPESKTIIEAVEEYLTLEGLRLIYERN 253  
 RESULT 9

AAU91151  
ID AAU91151 standard; Protein; 265 AA.

AC AAU91151;

DT 05-JUN-2002 (first entry)

XX Streptomyces coelicolor pantothenate kinase Coax.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Streptomyces coelicolor.

OS WO200216601-A2.

PN 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

PA Yocum RR, Patterson TA;

PI WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

XX Claim 10; Page 69-70; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX Sequence 265 AA;

Query Match 22.6%; Score 295; DB 23; Length 265;

Best Local Similarity 32.9%; Pred. No. 1e-25; Mismatches 54; Gaps 13;

Matches 94; Conservative 42; Indels 96; Indels 54; Gaps 13;

QY 1 MRLVVDIGNSTTLAFTGDEEPSVE---SVPSALFADSSST-----MREVFQGNMAR 48

Db 1 MLLTIDVGNTHTVLGLFPG--EDIVEHNRISTDSRRTADELAVLQGLMGHPLIGD--- 55

QY 49 KHGEP-QAIAICSVVPSAT-----AVGSALLES--LFSVPVLTICCKLRFPFR 93

Db 56 ELGGDIDGAIACATVPSVYLHELREVTTRYGDPVAVLVEPVGKTVGPILT----- 105

QY 94 LDYATPHFPFGADRLCALCAWSRHLFSEKPIAVDICTAITFDVLTGNYRGGLIMPGIDM 153

Db 106 -DH--PREVGADRIINAVAVELYG-GPAIVVDFGTATFDFAVSARGETVIGVAPGIEI 161

QY 154 MAGALHSRTAQLPQVRIDPRLSLGRSTTEICKSGVFWGVKQIGGLVDALRGDLVDFG 213

Db 162 SVEALGVKGAQLRIEVARPSVIGKNTVEAMQSGIVYFGAQVDGVVNRARELADD-- 219

QY 214 ESTVEVITGNSRIIVPEIGVPSVIDE---LAVLRGSDLLLRNMN 256

Db 220 PDDVTVTATGGLAPMV---LGESSVIDBHEPWLTLMLGLRLVYERNV 262

RESULT 10

AAU91152  
ID AAU91152 standard; Protein; 272 AA.

AC AAU91152;

DT 05-JUN-2002 (first entry)

XX Mycobacterium tuberculosis pantothenate kinase Coax.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Mycobacterium tuberculosis.

PN WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

PA Yocum RR, Patterson TA;

PI WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

XX Claim 10; Page 70-71; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX Sequence 272 AA;

Query Match 21.1%; Score 274.5; DB 23; Length 272;

Best Local Similarity 29.3%; Pred. No. 2.6e-23;

Matches 77; Conservative 52; Mismatches 125; Indels 9; Gaps 5;

QY 1 MRLVVDIGNSTTLAFTGDEEPS--VES---VPSALFADSSSTMRVFGNMARKHGEPO 54

Db 1 MLLAIDVRNTHTVVGLSGKHEKAKVVOQWRITSEVTADELAL-TIDGLIGDSERLT 59

QY 55 AIAICSVPSATAVGSALLESLF-SVPVLTICCKLRFPFRLDYATPHFPFGADRLCALCAWS 113

Db 60 GPAALSTVPSVLEHVRIMLDQYWPSPVPHVLEPGVTGIPLLVDNPKVEGADRLVNCIAA 119

QY 114 RHLFSEKPIAVDICTAITFDVLTGNYRGGLIMPGIDMAGALHSRTAQLPQVRIDRP 173

Db 120 YDRF-RKAAIVVDFGSSICVDVVSAGFELGGAIPGVQVSSDAAAARSALRVELARP 178

QY 174 ESSLGRSTTEICKSGVFWGVKQIGGLVDALRGDLVDFGSEVVEVITGNSRIIVPEI 233

Db 179 RSVVGKNTVECHQAGVAFGLVDGLVGRIRDVSFGSVYDHDAIVATGHTAPLLPEL 238

QY 234 GPVSVIDEALVLRGSDLLLRNMN 256

Db 239 HTVDHYDQHLTLQGLRLVFERNL 261



## RESULT 11

AAU91163  
\* ID AAU91163 standard; Protein; 233 AA.  
AC AAU91163;  
XX  
XX  
DT 05-JUN-2002 (first entry)  
XX  
XX  
DE Pantothenate kinase (Coax) #1.  
XX  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
XX  
OS Bacillus subtilis.  
XX  
XX  
PN WO200216601-A2.  
XX  
XX  
D 28-FEB-2002.  
X  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX  
PI Yocum RR, Patterson TA;  
XX  
XX  
DR WPI; 2002-269358/31.  
DR N-PSDB; ABK54169.  
XX  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
XX  
PS Disclosure; Page 81-82; 128pp; English.

XX  
XX  
CC The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

SQ Sequence 233 AA;

Query Match 21.0%; Score 273.5; DB 23; Length 233;  
Best Local Similarity 28.5%; Pred. No. 2.7e-23;  
Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;  
QY 1 MRLVVDIGNTSTLAIF-----TGDEEPSVESVPSALFADSSSMREVF 43  
DB 1 MLLVIDGNTVTLGVHDGKLEVHWRIETSRHKTDEFGMI-----LRSLF 47  
QY 44 GNMARKEGPOAIAICSVPSATAVGSALLSFLSPVLTICCKLRFPPRLDYAPHTFG 103  
DB 48 DHSGLMFEQIDGIITSSVPPIMFALERMCTKYPHIEPQIVGPGMKTGLNIKYNPKEVG 107  
QY 104 ADRLALCAWSRHLSEKPIAVDGTALTDFVLDVGNVGRGLIMPIDMAGALHSRTA 163  
DB 108 ADRIVNAVAALHLYG-NPLIVVDFGTATTCYIDENKQYMGGAIAPIGTISTEALYSRAA 166  
QY 164 QLPQVRIDRPSLLGRSTTECIKSGVFWGVKQIGGLVDAIRGDLVRD 211  
DB 167 KLPRIETRPDNIIGKNVTSAMQSGILFGYGVQVEGIVKRWKQAKQD 214

## RESULT 12

AAG81225  
ID AAG81225 standard; Protein; 272 AA.  
XX  
AC AAG81225;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 276.  
XX  
KW Drug target; growth; organism viability; characterisation.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200135317-A1.  
XX  
PD 17-MAY-2001.  
XX  
PF 13-NOV-2000; 2000WO-US31152.  
XX  
PR 12-NOV-1999; 99US-0165086.  
PR 12-NOV-1999; 99US-0165124.  
PR 01-FEB-2000; 2000US-0179531.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Eisenberg D; Rotstein SH, Marcotte EM;  
PI  
DR WPI; 2001-329193/34.  
DR N-PSDB; AAH52076.

Identifying nucleotide or polypeptide sequence for use as drug target,  
involves providing algorithm that analyzes a functional relationship  
between nucleotide or polypeptide sequences, and comparing the  
sequences

PS Disclosure; Page 188; 207pp; English.

XX  
XX  
CC This invention relates to a method for identifying a nucleotide or  
CC polypeptide sequence that may be a drug target, or essential for growth  
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium  
CC tuberculosis proteins which are potential drug targets. The DNA and  
CC protein sequences are used to illustrate the method of the invention. The  
CC method involves providing an unknown nucleotide or polypeptide sequences,  
CC and comparing it to a number of sequences along with at least one  
CC algorithm capable of analysing a functional relationship between  
CC nucleotide and polypeptide sequences. The method is useful for  
CC characterising the function of nucleic acids and polypeptides that may be  
CC useful as a target for a drug or essential for the growth or viability of  
CC an organism.

SQ Sequence 272 AA;

Query Match 20.8%; Score 271.5; DB 22; Length 272;  
Best Local Similarity 29.1%; Pred. No. 5.9e-23;  
Matches 76; Conservative 52; Mismatches 124; Indels 9; Gaps 5;  
QY 3 LVVDIGNTSTLAIFGTDEEPS--VES---VPSALFADSSSTREVEFGNMARKHGEPOAI 56  
DB 3 LAIDVRNTHTVVGLLSGMKEHAKVVOQWRIETSEVTADALAL-TIDGLIGESERTLTGT 61  
QY 57 AICSVVPSATAVGSALLSFL-SVPVLTICCKLRFPPRLDYATPHTFGADRLALCAWSRH 115  
DB 62 AALSTVPSVLEHVRIMLDQWSPVPHVLTIEPVRTGTPLLVDNPKVEGADRVNCLAAVD 121  
QY 116 LFSKEPVIADVTGTAITFDVLTGVNVRGLIMPIDMAGALHSRTAQIPQVRIDRPS 175  
DB 122 RF-RKAAIVVDFGSSICVDVWSAKGFEFLGATAPGVQVSSDAAAARSALRRVELARPS 180  
QY 176 LLGRSTTECIKSGVFWGVKQIGGLVDAIRGDLVRDVGESTVEVITVGGNSRIIVPEIGP 235  
DB 181 VYGKNTVECMQAGAVEFGAGLDGLVGRINREDSVGSVDHDAIVATGHTAPLLLP 240



Db 215 MKVIATGG 222 :||| |||  
QY 228 IIVPEIGPVSVIDEI LVLRLGSDLLLRMN 255 :||| :|||  
Db 228 LFEGATDSIDHFDSDLTIRGLLEIYARN 255 :||| :|||

Search completed: June 24, 2003, 22:08:36  
Job time : 30.091 secs

RESULT 15  
AAU91173  
ID AAU91173 standard; Protein; 260 AA.

XX AC AAU91173;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #11.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Caulobacter crescentus.

XX WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX DR WPI; 2002-269358/31.

XX DR N-PSDB; ABK54194.

XX PT Identifying potential antibiotic or antimicrobial agent, comprises  
XX PT contacting composition comprising pantothenate kinase (Coax) protein  
XX PT with test compound and identifying inhibitor of the Coax protein  
XX PS Claim 10; Page 102-103; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic  
XX CC comprising contacting an assay composition comprising a pantothenate  
XX CC kinase (Coax) protein with a test compound, and determining the ability  
XX CC of the test compound to inhibit the activity of the Coax protein, an  
XX CC essential enzyme for the production of coenzyme A. Coax protein is a  
XX CC valuable target for identifying bactericidal compounds. Coax modulating  
XX CC agents can be used in an infectious animal model to determine the  
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This  
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX CC described in the invention.

XX SQ Sequence 260 AA;

Query Match 20.0%; Score 261; DB 23; Length 260;  
Best Local Similarity 31.0%; Pred No. 9 3e-22;  
Matches 83; Conservative 41; Mismatches 118; Indels 26; Gaps 9;

QY 1 MRLVVDIGNSTTLAFTGDE-----EPSVESVPSA----LFPADSTMRVEFGNMARKHG 51  
Db 1 MLLATEQGNNTMFADHDSWVAQWRGATSTRTADEYVWLSQLLSMOGLGFRA---- 56  
QY 52 EPQATAICSVPSATAVSALLESFSPVLTII--CCKLRFPFRDLDPHTFFGADRLAL 109  
Db 57 -IDAVIISVVPQSFNLRNLSRRYFNVEPLVGENAKLIGIDVRIE--KPSEAGADRLVN 113  
QY 110 CAWSRHLSEKPEVIAVDIGTATFDVLDTVGNVYRGGLMPGIDMMAGALHSRTAQLPOVR 169  
Db 114 AIGAAWY-PGLVVDISGTATFDIVADGAFEGGIAPGINSQALHEAAKLPRIA 172  
QY 170 IDRP--ESILGRSTTECKSGVFWGVKQIGGLVDVIRDFGSESTVEIVTGNSR 227  
Db 173 IORPAGNRIVGTDVTSAMQSGVFWGYSILIEGLVARIKAEK----GE-PMTVIATGGVAS 227

**THIS PAGE BLANK (USPTO)**







## RESULT 8

E97293  
Probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridia  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: E97293  
R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 193, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97293  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
Gene: CAC3200  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 20.6%; Score 268; DB 2; Length 273;  
Best Local Similarity 28.4%; Pred. No. 1.5e-16;  
Matches 74; Conservative 59; Mismatches 114; Indels 14; Gaps 6;

QY 3 LVVDIGNSTTLAIF-----TGDEPSVESVPSALFADSTWREVFVGNMARKHGEP---Q 54  
DB 14 LVLDVGNNTVILGIYNDTKLTAEWRLSTDLVRS---ADEYGI-QVNNLFQDKLDPTLVE 69  
QY 55 AIAICSVVPSATVAVGSALLSFVPLTICCKLRPPFLDYATPHFTGADRILALCAWSR 114  
DB 70 GVILSVVPNIMYSLHEHMRKFKINPLVVGIGIKTGINIKNDPNKEVGADRI-VNAVAA 128  
QY 115 HLFSEKPVAVNDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSRTAQLPQVRIDRPE 174  
DB 129 HEIYRSLLIIDFGTATTCAVRENGDYLGAICPGIKVSSEALPEKAALPRVELIKPA 188  
QY 175 SILGRSTTECIKSGVFWGVKQIGGLVDARGLDVRDFGSEVVEIVTGGNRIIVPEIG 234  
DB 189 YAICANTISSIGSIVGIVGQVRYVERMKEL-QEGERKPLVAVTGGALKLISEAK 247  
QY 235 PVSVIDELAVLRGSDLLLRMN 255  
DB 248 NVDVINGPFLTEGLIIVKKN 268

## RESULT 9

7489  
anscription activator, probable Baf family [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87489  
R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87489  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1935  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 20.0%; Score 261; DB 2; Length 261;  
Best Local Similarity 31.0%; Pred. No. 6e-16;  
Matches 83; Conservative 41; Mismatches 118; Indels 26; Gaps 9;

QY 1 MRLVVDIGNSTTLAIFTGDB-----EPSVESVPSA-----LFADSSSTMREVFVGNMARKHG 51  
DB 2 MLALAEQGNVNTFAHDGASVWAQWRSATSESTFTADEYVYVWLSQQLSMQGLGFRA----- 57  
QY 52 EPOAIAICSVVPSATVAVGSALLSFVSPVLTII--CCKLRFPFLDYATPHFTGADRILAL 109  
DB 58 -IDAVIISVVPSQIFNLNLSRRYFNVEPLVIGENAKLIGDIVRIE--KPSEAGADRLVN 114  
QY 110 CANSRHLFSEKPVAVNDIGTAITFDVLDVTGNYRGGLIMPIDMAGALHSRTAQLPOVR 169  
DB 115 AIGAAVMY-PGLVWIDSGTATTFDIVAADGAFEGGIIAFGINLSMOLHEAAKLPRIA 173  
QY 170 IDRP--ESLLGRSTTECIKSGVFWGVKQIGGLVDARGLDVRDFGSEVVEIVTGGNSR 227  
DB 174 IQRPAGNRIVGTDTVSAMQSGVFWGYSLSIEGLVARIAKER----GE-PMTVIATGGVAS 228  
QY 228 IIVPEIGPVSVIDEAVLRGSDLLLRMN 255  
DB 229 LPFEGATSDIDHFDSDLTIRGLLEIYRN 256

## RESULT 10

E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A73250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WHI>  
A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 19.5%; Score 253.5; DB 2; Length 262;  
Best Local Similarity 27.8%; Pred. No. 2.8e-15;  
Matches 74; Conservative 46; Mismatches 103; Indels 43; Gaps 7;

QY 3 LVVDIGNSTTLAIFTGDEPSVESVPSALFADSS-----TMREVFEGN--MARKH 50  
DB 7 LAVDIGNTTVLGL-----ADAGALTHTWRTNREMLPDDDLALQLH 49  
QY 51 G-----EPOAIAICSVVPSATVAVGSALLSFVSPVLTICCKLRFPFLDYATPHFT 102  
DB 50 GLFTLAGAIPRAAVLSSVAPPVGENYALAKRHFMDAFAVSAENLPDVTVELDTPGSV 109  
QY 103 GADRIA-LCAWSRHLFSEKPVIAVDIGTAITFDVLTGNYRGGLIMPIDMAGALHSR 161  
DB 110 GADRLCNLFGEAKYGLGLDYAVVVDFTGSTNFDVGRGRFLGGLIATGAQVSADALFAR 169  
QY 162 TAQLPQVRIDRPESELGRSTTECIKSGVFWGVKQIGGLVDARGLDVRDFGSEVVEIV 221  
DB 170 AAKLPRLITLQAPETAIGKNTVHALQSLGVFGYAEVMDGLRRIRAEPL---GEAV---AVA 224  
QY 222 TGGNSRIIVPEIGPVSVIDEAVLRG 247  
DB 225 TGFSPRTVQICQEIYDDETILRG 250

## RESULT 11

D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima



Db 40 PKLKALGISVKQS-FSEKVRGKIPKIKFLAKENFIQVDYKTPETLGTDRVAL-AYSAKK 97  
 QY 117 FSEKPVIAVDIGTAIFDVLDTVGNTRGGGLIMPICIDMMAGALHSRTAQLPOVRIDRPESL 176  
 Db 98 FYGKNVVVISAGTAIVIDLVL-LSGKPKGGFTTGLGKKLALSDLAEGSIPEFFEEVF 156  
 QY 177 LGRSTTECKSGVFMGVKQIGGLDAIRGDLVRDFGSGESVEVITVCGNSRIIVPEIGPV 236  
 Db 157 LGRSTRECVLGGAYRESTEPFKSTLKLWRKVKFRKF-----KVITGEGKY----FSKF 207  
 QY 237 SVIDELAVLRGSDLL 252  
 Db 208 GIYDPLLVHGRNLL 223  
  
 RESULT 13  
 AI2292  
 hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)  
 C;Species: Nostoc sp.  
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C;Accession: AI2292  
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobact  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AI2292  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-276 <GR>  
 A;Cross-references: KUB>BA000019; PIDN:BA075595.1; PID:gl7133030; GSPDB:GN00179  
 A;Experimental source: strain PCC 7120  
 C;Genetics:  
 A;Gene: alr3896  
  
 Query Match 16.7%; Score 218; DB 2; Length 276;  
 Best Local Similarity 27.2%; Pred. No. 4.8e-12;  
 Matches 79; Conservative 46; Mismatches 93; Indels 72; Gaps 13;  
 OV 3 LVVDIGNSTTTLAFTGDEEPSVESVPSALFADSSTMREVFGNMKRHGE----- 52

```

15 LALEIGNSRUHWALFWGE-----SLEF7WDYILPESVLIQQL-GN-----GETKLEVGSEER 63
QY -----PQAIACISVVP SATAVGSALLESLFSVPVLTICCKLRFPFRLDYAT 98
Db EIEFFTEFLPPACPULPLFIASVVPQOT-----VLWENYLNVRTILD-----QIPLNNIYP- 117
QY PHTFGADRALCAWSRHLSSEKVPVIAVDIGTATFDVLTGVNRYGGLIMPIDMAGAL 158
118 --FLGIDR-ALAUWAGMSGFPVLVIDAGTALTFTTAADGCKNLVGGAILPEVGLQFASL 174
QY HSRTAQLPQVRIDRPSLGR-----STTECIKSGVFWGVKQIGGLVDAIRGDLVRDFQES 215
Db GQQTGQLPQVEMEAIKSLPRFALNTTEAIQSGVIYTL-----IAGM-----RDFTEE 222
QY 216 TVEVITVG-----GNSRIIV-----PEIGPVSVIDELAVLRGSDLLL 252
223 WLSFLFDPGKVAIKGGORILLNLYLQALYPDLAARLIVEPNLIFWGMQTIIV 272

```

D1J326  
conserved hypothetical protein TP0431 - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D1J326  
R:Raser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback,  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete  
A:Reference number: A71250; UID:98332770; PMID:9665876

A:Accession: D71326

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <COL>

A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0431

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 15.2%; Score 197.5; DB 2; Length 273;  
Best Local Similarity 25.7%; Pred. No. 3.3e-10;  
Matches 71; Conservative 43; Mismatches 117; Indels 45; Gaps 8;

QY 3 LVVDIGNSTTIAITGDEEPSVESPSALFADSTMTREVF--GNMARKHGEPOAIAICS 60

Db 2 LIIDVGNSHVVGII--OGENGGRV-----CVRELFRLAPDARKTODEYSLLIHA 48

QY 61 VVPSATAVGSALLESLFSVPVLTICCKL-----RPFRLDYAT 98

49 LCERA-GVGRSLRADAFISSVVPVLTITKTADAVAQISGVQPVVFGPWAYEHLFVRIPEPV 107

99 PHTFGADRLALCAWHRHLFSEKPVIAVDIGTAITFDVLTGVNRYRGLIMPIDMMAGAL 158

Db 108 RAEIGTD-LVANAVAAVYHFRSACVVDGCTALTFTAVDGTGLIQVALAPGLRTAVQSL 166

QY 159 HSRTAQLPQVRIDRPESLLGRSTTECIKSGVFWGVVQIGGLVDAIRGDLVDRDFGESIVE 218

Db 167 HGTGRLPLVPLALPDSVLGKDTTHAVAGVGVGTFLVIRAMIAQCCKEL-----GCRCA 221

QY 219 VIVTGNRIIVPEIGPVSVIDEALVLRGSDLLLRM 254

Db 222 AVITGSLRSLFSEV-DFPPIDAQLTSLGLAHIALRL 256

#### RESULT 15

F70165

conserved hypothetical protein BB0527 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C:Accession: F70165

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kervavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390: 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; PMID:98065943; PMID:9403685

A:Accession: F70165

Status: preliminary; nucleic acid sequence not shown; translation not shown

Molecule type: DNA

A:Residues: 1-262 <KLE>

A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g268843

A:Experimental source: strain B31

Query Match 12.8%; Score 166.5; DB 2; Length 262;  
Best Local Similarity 24.5%; Pred. No. 1.9e-07;  
Matches 60; Conservative 45; Mismatches 101; Indels 39; Gaps 7;

QY 3 LVVDIGNSTTIAITGDEEPSVESPSALFADSTMTREVFGNMARKHGEPOA-----55

Db 9 LIIDIGNTSIAFLFKDQ-----VNLFIKMT-----NMLRYDEVYSPFEENFD 54

QY 56 -----IAICSVVPSATAVGSALLESLFSVPVLTICCKLRF-----PFLDYATPHTFGAD 105

Db 55 FNVNKFVIVSSVPLNETFNKVFIFSFKIKPLFGDLNIDLTENPKSD---KFLGSD 111

QY 106 RLALCAWHRHLFSEKPVIAVDIGTAITFDVLTGVNRYRGLIMPIDMMAGALHSRTAQL 165

Db 112 VFANLVAENISFENVLVDLGFACITFAVSRODGIIGLINSGLINFNLSLDNAYLI 171

QY 166 PQVRIDRPESLLGRSTTECIKSGVFWGVVQIGGLVDAIRGDLVDRDFGESIVEIVTGN 225

Db 172 KKFPITPNNLLERTTSGSVNSGLFY----QYKYLIESVYRD-IKQMYKKKFNLIITGN 226

QY 226 SRIIV 230

Db 227 ADLIL 231

Search completed: June 24, 2003, 22:19:18  
Job time: 11.6025 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.16979 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-53

Perfect score: 1303

Sequence: 1 MRLVVDIGNTSTTIAFTGDD.....VIDELAVLGRSGLLRLMMP 257

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	273.5	21.0	233	1 YACB_BACSU	P37564 bacillus su
2	136.5	10.5	267	1 BAF_BORPE	Q45338 bordetella
3	92.5	7.1	747	1 PURL_DEIRA	Q9rxrt4 deinococcus
4	91.5	7.0	554	1 SECD_RHOCA	Q33517 rhodobacter
5	88.5	6.8	781	1 DPOL_ARCFU	Q29753 archaeoglob
6	88.5	6.8	786	1 AAC_ACTUT	P29958 actinoplan
7	85.5	6.6	592	1 HIS5_ARATH	Q9sz30 arabidopsis
8	85.5	6.6	835	1 OBP_VZVD	P09299 varicella-z
9	85	6.5	351	1 MREB_HAEIN	P44474 haemophilus
10	82.5	6.3	810	1 SYFB_SYNY3	P74296 synechocyst
11	82	6.3	351	1 MREB_PASMU	Q916a3 pasteurella
12	81	6.2	260	1 HGDC_ACFIE	P11568 acidaminoco
13	81	6.2	1106	1 ACILY_CAEEL	P53585 caenorhabdi
14	80.5	6.2	384	1 NIFS_RHOCA	Q07177 rhodobacter
15	80	6.1	331	1 PGFB_RAT	Q08603 rattus norv
16	80	6.1	488	1 BTBL_MOUSE	P58544 mus muscucu
17	79.5	6.1	524	1 AMPA_AERPE	Q9y935 aeropyrum p
18	79	6.1	337	1 MREB_BACSU	Q01465 bacillus su
19	79	6.1	509	1 YMCB_BACSU	Q01778 bacillus su
20	79	6.1	1693	1 POLN_HEVBU	P29324 hepatitis e
21	79	6.1	1693	1 POLN_HEVMI	Q04610 hepatitis e
22	79	6.1	2021	1 OMPA_RICCN	Q52657 rickettsia
23	78.5	6.0	296	1 NPL_ECOLI	P06995 escherichia
24	78.5	6.0	533	1 PERC_ASPNG	P33295 aspergillus
25	78	6.0	373	1 PRXC_CALFU	P04963 caldariomyc
26	78	6.0	444	1 YOYA_BACSU	Q01862 bacillus su
27	78	6.0	495	1 AMPA_PSEAE	O68822 pseudomonas
28	78	6.0	575	1 SMFL_YEAST	P38925 saccharomyc
29	78	6.0	634	1 DNAR_HAEIN	P43736 haemophilus
30	78	6.0	760	1 MEPE_MYCLE	O05564 mycobacteri
31	77.5	5.9	528	1 UDPE_NPVMB	Q83140 namestra br
32	77.5	5.9	1403	1 VG22_HSVII	Q00105 ictaluriid h
33	77.5	5.9	4367	1 DYHC_NEUCR	P45443 neurospora

34	77	5.9	288	1 BLC6_VIBCH	P81781 vibrio chol
35	77	5.9	338	1 PSCI_MYCTU	P95303 mycobacteri
36	77	5.9	428	1 BCH2_RHOCA	P26171 rhodobacter
37	77	5.9	799	1 YDDU_ECOLI	P76129 escherichia
38	77	5.9	1037	1 YOJ8_YEAST	Q12496 saccharomyc
39	76.5	5.9	278	1 CYST_SYNP7	P27367 synechococc
40	76.5	5.9	342	1 HEMA_RICPR	Q92c84 rickettsia
41	76.5	5.9	463	1 VL2_HPVA5	P36761 human papil
42	76.5	5.9	916	1 PMAL_AJECA	Q07421 ajelomyces
43	76	5.8	360	1 BUK_ENTFA	Q9rps7 enterococcu
44	76	5.8	591	1 KPY2_SYNY3	P73534 synechocyst
45	76	5.8	1100	1 TRAA_AGR75	Q44349 agrobacteri

#### ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Broutlet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Deniset F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D., N.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche M., Rose M., Sadale F.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokina A., Tacconi E., Takagi T., Takahashi H., Takemura K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,  
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26185; BAA05305.1; -
DR EMBL; Z99104; CAB11846.1; -
DR Subtilisin; BG10133; yacB.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
DR TIGRams; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match 21.0%; Score 273.5; DB 1; Length 233;
Best Local Similarity 28.5%; Pred. No. 4.7e-17;
Matches 65; Conservative 42; Mismatches 90; Indels 31; Gaps 3;

1 MRLVVDIGNSTTIAIF-----TGDEPSVESVPSALFADSTSTREVF 43
1 MLLVVDIGNSTTIAIF-----TGDEPSVESVPSALFADSTSTREVF 47
44 GNMARKHGEPOQAICSVVPSATVGSALLESFVPLTICCKLRPFRLDYATHTFG 103
48 DSHGLMFEQIDGIIISVPPIMFALERMCTKYFIEPQIVGPMGTGLNIKYDNPKEVG 107
104 ADRLALCAWSRHLSFKPVIADIGTAITFDVLTGVNRYGGLIMPGIDMAGALHSRTA 163
108 ADRIVNAVAAILHYG-NPLIVDFGTATTCYIDENKQYMGGAIGTITSTALYSRAA 166
164 QLPQVRIDRPESLLGRSTTECIKSGVFWGVYKQIGGLVDAIRGDLVRD 211
167 KLPIETITRPDNIIGNTSVMSQGLIFGVYGVQVEGIVKRMKQWQAKD 214

RESULT 2
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45337;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
Bordetella.
NCBI_TaxID=520;
[1]
SEQUENCE FROM N.A.
RC STRAIN-BP504;
RX MEDLINE=9532323; PubMed=7601846;
RA Deshaizer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli."
RL J. Bacteriol. 177:3801-3807(1995).
[2]
SEQUENCE OF 1-38 FROM N.A.
RC STRAIN-BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bira homolog in Bordetella pertussis."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 239-267 FROM N.A.
RC STRAIN-BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus

```

```

RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis."
RL Mol. Microbiol. 19:37-52(1996).
CC -1- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12020; AAA75361.1; -
DR EMBL; AF016461; AAC68834.1; -
DR EMBL; X90711; CAA62242.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc.factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match 10.5%; Score 136.5; DB 1; Length 267;
Best Local Similarity 26.2%; Pred. No. 6.8e-05;
Matches 62; Conservative 31; Mismatches 113; Indels 31; Gaps 9;

1 MRLVVDIGNSTTIAIFTGDEPSVESVPSALFADSTSTREVFGNMARKHGEPO 54
1 MLLVVDIGNSTTIAIFTGDEPSVESVPSALFADSTSTREVFGNMARKHGEPO 59
55 --AIAICSVVPSATVGSALLESFVPLTICCKLRPFRLDYATHTFGADRLA--LC 110
60 VAGLARGEATAATLRAGGCCDIRWLRAQP-----LAMGLRNGYRNPDLQAGRWACMVG 112
111 AWSRHLSFKPVIADIGTAITFDVLTGVNRYGGLIMPGIDMAGALHSRTAQLP--Q 167
113 VLARQSVHPPLLVASFGTATLTGTGPDNVFPGGLILPGPMARGALAYTAHLPLADG 172
168 VRIDRPESLLGRSTTECIKSGVFWGVYKQIGGLVDAIRGDLVRDVGSTEVIVTGG 224
173 LVADYP-----IDTHQAIASSI---AAQAAGAIYVQWLAGRQR-YGQAP-EIYVAGG 219

RESULT 3
ID PURL_DEIRA STANDARD; PRT; 747 AA.
AC Q9RXT4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
GN PURL OR DR0222.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Otterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-

```

CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-  
 CC (formamido)-N(1)-(5-phospho-D-ribosyl)acetamide + L-glutamate.  
 CC -!- PATHWAY: De novo purine biosynthesis; fourth step.  
 CC -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURQ AND PURL.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE001884; AAF09808.1;  
 CC TIGR; DR0222;  
 CC InterPro; IPR000728; AIRS-related.  
 CC Pfam; PF00586; AIRS; 2.  
 CC Purine biosynthesis; Ligase; ATP-binding; Complete proteome.  
 CC NP\_BIND 103 114  
 CC SEQUENCE 747 AA; 78366 MW; CBI69EEC6E063724 CRC64;  
 CC  
 CC Query Match 7.1%; Score 92.5; DB 1; Length 747;  
 CC Best Local Similarity 22.1%; Pred. No. 1.7;  
 CC Matches 64; Conservative 38; Mismatches 102; Indels 85; Gaps 14;  
 CC  
 CC QY 4 VVDIGNVSTTLAIGTDEEPS-VESVPSALPADSTMTREVEFGNMAK-----KHGPPQ 54  
 CC DB 76 VVDIGDGVAFKESHNPSPAVEPQGAATGVGILNIDFAMGARPAVLDLSLRFNGPD 135  
 CC QY 55 A-----IAICSVVPSATAGVSALESLSFVPLTICCKLRFFFLDYATPHFGADRLALC 110  
 CC DB 136 SPRTFLVGVGDIAGHYGNAI-----GVP--TVGGEVTF----- 168  
 CC QY 111 AWSRHLSEKPVIAV-----DIGTATFDVLDTVGNRGLIMPIDM 153  
 CC DB 169 ---HPSQENPLNVNMGALLRHEDLATGTMGVEGNQIVYVSGTGRDGLGAVFSSADL 225  
 CC QY 154 MAGALHSRTAQLPOVRIDRP--ESLLGRSTTECTKSGVFWGVKQIG--GLVD-----AI 204  
 CC DB 226 SAAS-----QADRPVAVQDPPMEKLLLEATLQAGLVAG-VQDMGAAGLVSSCEMAY 280  
 CC QY 205 RGLD-----VRDEGSTVEVITGGNSR-IIVPEIG-PVSVIDEVLAVLR 246  
 CC DB 281 RASLGITWDLDKVPTREGMVPMELCUSEOERMILVPPVQKEALHDL 329  
 CC  
 CC RESULT 4  
 CC SECD\_RHOCA STANDARD; PRT; 554 AA.  
 CC ID DPOL\_ARCFU  
 CC AC 03517;  
 CC DT 30-MAY-2000 (Rel. 39, Created)  
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 CC DE Protein-export membrane protein secD.  
 CC GN SECD.  
 CC OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
 CC OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
 CC OC Rhodobacter.  
 CC OX NCBI\_TaxID=1061;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=SB1003 / St Louis;  
 CC RX MEDLINE=97318920; PubMed=9175857;  
 CC RA Goldman B.S., Beckman D.L., Ball A., Monika E.M., Gabbert K.K.,  
 CC RA Kranz R.G.;  
 CC RT "Molecular and immunological analysis of an ABC transporter complex  
 CC RT required for cytochrome c biogenesis."  
 CC RL J. Mol. Biol. 268:724-738(1997).  
 CC CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).  
 CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS

CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH  
 CC (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U69979; AAB62801.1;  
 CC InterPro; IPR003335; SecD\_SecF.  
 CC Pfam; PF02355; SecD\_SecF; 1.  
 CC TIGRFAMS; TIGR00916; 2A0604s01; 1.  
 CC TIGRFAMS; TIGR01129; secD; 1.  
 CC KW Protein transport; Translocation; Transmembrane; Membrane.  
 CC FT TRANSMEM 10 30 POTENTIAL.  
 CC FT TRANSMEM 392 412 POTENTIAL.  
 CC FT TRANSMEM 435 455 POTENTIAL.  
 CC FT TRANSMEM 491 511 POTENTIAL.  
 CC FT TRANSMEM 516 536 POTENTIAL.  
 CC SQ SEQUENCE 554 AA; 58943 MW; DF2CBEEBA9F69EDF CRC64;  
 CC  
 CC Query Match 7.0%; Score 91.5; DB 1; Length 554;  
 CC Best Local Similarity 24.5%; Pred. No. 1.5;  
 CC Matches 67; Conservative 39; Mismatches 111; Indels 57; Gaps 16;  
 CC  
 CC QY 16 IFTGTE-----EPSVE--SVPSALFADSTMTREVEFGNMAK--CEPOAIAICSVVPSATAV 68  
 CC DB 278 VVTGDDLTDRPTTDDNGAPAVSPFNWSGARAFGYTAGHGEPPAIVLDGKVISAPTI 337  
 CC QY 69 -----GSALLESLFSVPVLT-ICCKLR---PFRLDYATPHT-----FGADR----- 106  
 CC DB 338 QAHIAGGSGIITGRFSIEEATDLALLRAGALPAGMTFLEERTIGFELGADSVKAGMVAS 397  
 CC QY 107 -----LALCAW-----SRHLSEKPVIAVDIGTATFDVLDVGNRGLIMPIDMAGAL- 158  
 CC DB 338 VIGFVAVVAYMIASYGLGFSSVALFINAFVAVGAIG--GTMTLPGI---AGIVL 451  
 CC QY 139 ---HSRTAQLPOVRIDRPESLLGRSTTECTKSGVFWGVKQIGLVDAIRGLDVRDFGES 215  
 CC DB 452 TIGTSVDANVLIERMRREIRSGKSPVRAIE-----LGFDKAMSAIIDA-----NVSFLSS 503  
 CC QY 216 TVEVITVGNSR--IIVPEIG-PVSVIDEVLAVLR 246  
 CC DB 504 AILFVLGAGPVRGFAVTMTIGTAASIFTAIWVR 537  
 CC  
 CC RESULT 5  
 CC DPOL\_ARCFU  
 CC ID DPOL\_ARCFU STANDARD; PRT; 781 AA.  
 CC AC 029753;  
 CC DT 15-JUL-1998 (Rel. 36, Created)  
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE DNA polymerase (EC 2.7.7.7).  
 CC GN POL OR POLB OR AF0497.  
 CC OS Archaeoglobus fulgidus.  
 CC OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 CC OC Archaeoglobaceae; Archaeoglobus.  
 CC OX NCBI\_TaxID=2234;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 CC RX MEDLINE=98049343; PubMed=9389475;  
 CC RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 CC RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 CC RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
 CC RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 CC RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,



amidotransferase/cyclase suppresses the histidine auxotrophy of a  
Saccharomyces cerevisiae his7 mutant.";  
FEBS Lett. 428:229-234(1998).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=cv. Columbia;  
MEDLINE=20083488; PubMed=10617198;  
Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
Pohl T., Dueterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
Harris B., Ansoorge W., Brandt P., Grivelli L.A., Rieger M.,  
Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
Kreiss M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
Langham S.-A., McCullagh B., Bilham L., Robben J.,  
Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
Braeken M., Weltjens I., Voet M., Bastiaens U., Aert R., Defoor E.,  
Weitzneger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
Clark L., Doggett J., Hall S., Kay M., Lennard N., McKay K., Mayes R.,  
Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
Neumann S., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,  
Massenot O., Quigley A., Vitale D., Liguori R., Piravandi E.,  
Schubel S., Hiller R., Clabaud G., Muendlein A., Felber R.,  
Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,  
Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
Heljnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
Sekhon M., Murray J., Sheel P., Cordes K., Abu-Threiden J.,  
Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
Granat S., Shohdy N., Hasegawa A., Rameed A., Lodhi M., Johnson A.,  
Chen E., Marra M., Martienssen R., McCombie W.R.;  
"Sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana.";  
Nature 402:769-777(1999).  
-1- FUNCTION: IGP5 catalyzes the conversion of PRPAR and glutamine to  
IGP, AICAR and glutamate. The glutamine amidotransferase domain  
provides the ammonia necessary to the cyclase domain to produce  
IGP and AICAR from PRPAR.  
-1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-  
ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-  
carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-  
aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.  
-1- PATHWAY: Histidine biosynthesis; fifth step.  
-1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISF  
FAMILY.  
-1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

-----
CC EMBL; AB006210; BAA28783.1; -
DR EMBL; AB016783; BAA32287.1; -
DR EMBL; AL035440; CAB36536.1; -
DR EMBL; AL161566; CAB79545.1; -
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR004651; HisF.
DR InterPro; IPR000570; His_biosynth.
DR Pfam; PF00111; GATase; 1.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRfam; TIGR00735; hisF; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW Lyase; Multifunctional enzyme; Chloroplast; Transist peptide.
FT TRANSIT ? 592
FT CHAIN ? 592
FT DOMAIN 61 263
FT AMIDOTRANSFERASE.
FT CYCLASE.
FT ACT_SITE 141 141
FT GATASE (BY SIMILARITY).
FT ACT_SITE 246 246
FT GATASE (BY SIMILARITY).
FT ACT_SITE 248 248
FT GATASE (BY SIMILARITY).
FT ACT_SITE 289 289
FT POTENTIAL.
FT ACT_SITE 447 447
FT POTENTIAL.
FT CONFLICT 110 110 A -> P (IN REF. 1).
FT CONFLICT 153 153 E -> Q (IN REF. 1).
FT CONFLICT 405 405 I -> M (IN REF. 1).
FT CONFLICT 413 413 S -> F (IN REF. 1).
FT CONFLICT 487 487 R -> Q (IN REF. 1).
FT CONFLICT 515 515 D -> N (IN REF. 1).
FT CONFLICT 557 576 KTNASAAALAAAGFFHKEVPI -> EDKRICRACCRHFFPPER
FT GYQS (IN REF. 1).
SQ SEQUENCE 592 AA; 64192 MW; A8D7F1699C6163CE CRC64;

Query Match 6.6%; Score 85.5; DB 1; Length 592;
Best Local Similarity 25.6%; Pred. No. 5.5;
Matches 41; Conservative 27; Mismatches 51; Indels 41; Gaps 8;

QY 27 SVPSALFADSTMRVFGNMARKEGPOAIAICSVVPSATAVGSALLESLSFVPLTICC 86
DB 27 SSPASLFSQKSI-----GNVRKFKPSRL-----SVRASSTSDSVVTLDDYGAGNVRISRN 79
QY 87 KLR---PPFRDYPATP-HTFGADRL-----ALCAWSRHLEFSEK 120
DB 80 ALRHGFSIK-DVQTPGDILNADRLIFPGVGAFAPAMVDNRTGMREALC---KYIENDR 135
QY 121 PVIAVDIGTATFDVLDTVGNRYRGLIMPGI----DMMAG 156
DB 136 PFLGICLGQLLFDSSSEENGPKVLGIVPGIVGREDASAG 175

RESULT 8
OBP_VZVD STANDARD; PRT; 835 AA.
AC P09299;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Replication origin binding protein (OBP).
GN 51.
OS Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_Taxid-10338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86306657; PubMed-3018124;
RA Davison A.J., Scott J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. Gen. Virol. 67:1759-1816(1986).
CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,

```

```

EHV-1 53, AND VZV 51.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X04370; CAA27933.1;
PIR: G27344; WZB551.
InterPro: IPR003450; Herpes_ori_bp.
Pfam: PF02399; Herpes_ori_bp; 1.
KW DNA replication; DNA-binding; ATP-binding.
FT NP_BIND 67 74
SQ SEQUENCE 835 AA; 94374 MW; A71F5877ACF386FB CRC64;

Query Match          6.6%; Score 85.5; DB 1; Length 835;
Best Local Similarity 22.1%; Pred. No. 8.2;
Matches 55; Conservative 30; Mismatches 93; Indels 71; Gaps 9;

QY 49 KHGEPOAIAICSV-VPSATAVGSALLESL-----FSVPVLTICCKLRFPFRL----- 94
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 53 KPGMSQTRPVTVVRAPMGSGKTTALLEWLOHAKADISVLVVSRRSFTQTLIQRFENDAG 112
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 95 -----DYATPHTF--GADRLALCANSHLSEKPIVANDIGTAITFDVLDVTGNYRGGLI 147
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 LSGFTYLTSETYIMGFKRLIVQLESLSRVSEADISYDV-----LI 154
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 148 MPGIDMAGALHSRAQLPQVRIDRPESLLGRSTTECKISGVFWGV--KQIGGLVDAIRG 206
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 LDEVMSVIGQLXPTWR-----RLSAVDSLLYLLNRCSQIIAMDATVNSQFIDLISGLRG 210
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 207 D-----LVRFGESTVEVIVTGGNSRIIPEIGFVSVIDEVLAVL 245
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 211 DENIHTIVCTYAGVFGSGRTCTILRDMGIDTLVRVTKRS-----PEHEDVRTIHQLRGT 264
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 RGSDLLLRM 254
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 FDELALRL 273
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
MREB_HAEIN
ID MREB_HAEIN STANDARD; PRT; 351 AA.
AC P44474;
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rod shape-determining protein mreB.
GN MREB OR HI0037.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney J., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Science 269:496-512(1995).
CC -!- FUNCTION: INVOLVED IN FORMATION OF THE ROD SHAPE OF THE CELL

```

```

(CY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE FTSA/MREB FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U32689; AAC21715.1; ALT_INIT.
TIGR: HI0037;
InterPro: IPR001023; Hsp70.
InterPro: IPR004753; MreB_Mrl.
DR PRODOM: PD000089; Hsp70; 1.
DR TIGRFAMS: TIGR00904; mreB; 1.
KW Cell shape; Complete proteome.
SQ SEQUENCE 351 AA; 37616 MW; D97CA0686403B279 CRC64;

Query Match          6.5%; Score 85; DB 1; Length 351;
Best Local Similarity 23.0%; Pred. No. 3.3;
Matches 50; Conservative 35; Mismatches 66; Indels 66; Gaps 10;

QY 52 EQQATAICSVPSATAVGSALLE--SLFSVPVLTICCKLRFPRLDYATPHTFGADRL- 107
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 144 EPMAAAGAKPLVSTAVGSMVIDIGGGTTEVAVISL-----NGIVYSSVRIGGRFD 196
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 108 -ALCAWSRHLFSE--KPV---IAVDIGTAITFDVLDVTGNYRGGLIMPGIDMMAGALHS 160
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 197 EALISVVRFTGSGVIGEPTAEIKQEGSAIQE---GDEIKEMEVHGHNLAEAPRS 251
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 RTAQLPQVRIDRPESLLGRSTTECKISGVFWGVVQIGGLVDAIRGLDYRFGESTVEV- 219
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 252 FT-----LTSRDVLEAIQ-----QPLNGIVAAVRALECECPHEAADI 290
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 ----IVTGGNSRIIPEIGFVSVIDEVLAVLGRSDDL 252
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 291 ERGMVLTGGG-----ALLRNIDILL 310
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
SYFB_SYNY3
ID SYFB_SYNY3 STANDARD; PRT; 810 AA.
AC P74296;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PHERS).
GN PHET OR SLL1553.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) -> AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 1.

```











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.9453 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-11

Perfect score: 1331

Sequence: 1 MNKPLLSLLIDIGNTSIAF.....HLTVEGVRIILGNSIDFKFVN 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	100.0	262	9	US-09-813-453A-11
2	282	21.2	256	9	US-09-813-453A-55
3	276.5	20.8	262	9	US-09-813-453A-45
4	275.5	20.7	254	9	US-09-813-453A-47
5	271.5	20.4	258	9	US-09-813-453A-2
6	266.5	20.0	255	9	US-09-813-453A-7
7	262	19.7	250	9	US-09-813-453A-3
8	256.5	19.3	258	9	US-09-813-453A-49
9	231.5	17.4	233	9	US-09-813-453A-17
10	203	15.3	246	9	US-09-813-453A-19
11	197.5	14.8	258	9	US-09-813-453A-6
12	190	14.3	219	9	US-09-813-453A-57
13	189.5	14.2	260	9	US-09-813-453A-51
14	181.5	13.6	265	9	US-09-813-453A-4
15	170	12.8	273	9	US-09-813-453A-10
16	166.5	12.5	257	9	US-09-813-453A-53
17	159	11.9	262	9	US-09-813-453A-8
18	141.5	10.6	229	9	US-09-813-453A-12
19	138	10.4	272	9	US-09-712-363-276

20	138	10.4	272	9	US-09-813-453A-5	Sequence 5, Appl
21	136	10.2	241	9	US-09-813-453A-63	Sequence 63, Appl
22	135	10.1	244	9	US-09-813-453A-41	Sequence 41, Appl
23	121.5	9.1	209	9	US-09-813-453A-21	Sequence 21, Appl
24	121.5	9.1	476	10	US-09-774-414-3	Sequence 3, Appl
25	111	8.3	212	9	US-09-813-453A-59	Sequence 59, Appl
26	100	7.5	592	9	US-09-813-453A-43	Sequence 43, Appl
27	99.5	7.5	410	10	US-09-845-335-2	Sequence 2, Appl
28	96	7.2	592	9	US-09-813-453A-22	Sequence 22, Appl
29	93.5	7.0	242	9	US-09-813-453A-65	Sequence 65, Appl
30	92	6.9	460	9	US-09-813-453A-39	Sequence 39, Appl
31	92	6.9	1161	9	US-10-170-102-4	Sequence 4, Appl
32	90	6.8	257	9	US-09-813-453A-13	Sequence 13, Appl
33	88.5	6.6	223	9	US-09-895-913A-74	Sequence 74, Appl
34	88.5	6.6	223	9	US-09-813-453A-14	Sequence 14, Appl
35	88.5	6.6	223	9	US-09-813-453A-67	Sequence 67, Appl
36	88.5	6.6	370	9	US-10-106-698-6268	Sequence 6268, Ap
37	86.5	6.5	385	10	US-09-815-242-5713	Sequence 5713, Ap
38	86.5	6.5	518	10	US-09-815-242-12473	Sequence 12473, A
39	85.5	6.4	284	9	US-10-284-986-1	Sequence 1, Appl
40	85.5	6.4	284	10	US-09-846-808-1	Sequence 1, Appl
41	84.5	6.3	381	9	US-09-870-759-84	Sequence 84, Appl
42	84.5	6.3	1536	9	US-10-092-880-2	Sequence 2, Appl
43	84	6.3	401	10	US-09-815-242-5506	Sequence 5506, Ap
44	84	6.3	401	10	US-09-815-242-12261	Sequence 12261, A
45	83.5	6.3	248	9	US-09-813-453A-20	Sequence 20, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-813-453A-11

; Sequence 11, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; PRIOR FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11:

; LENGTH: 262

; TYPE: PPT

; ORGANISM: Borrelia burgdorferi

US-09-813-453A-11

Query Match

Best Local Similarity 100.0%; Score 1331; DB 9; Length 262;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFTKMTNLMRLRYDEVYSFEEFNDFVNVKV 60

Db 1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFTKMTNLMRLRYDEVYSFEEFNDFVNVKV 60

Qy 61 FISSVVPILNETFKNVIESFFKIKPLFTGFDLNYDLTFTNPYKSDKFLGSDVFANLVAAI 120

Db 61 FISSVVPILNETFKNVIESFFKIKPLFTGFDLNYDLTFTNPYKSDKFLGSDVFANLVAAI 120

Qy 121 ENYGFENVLVVDLGTACTIFAVSRDGLGGILGGINSGLPINFNSLLDNAYLKKFPISTPN 180

Db 121 ENYGFENVLVVDLGTACTIFAVSRDGLGGILGGINSGLPINFNSLLDNAYLKKFPISTPN 180

Qy 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGGNADLILSLIETEFIF 240

Db 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGGNADLILSLIETEFIF 240

Db 181 NLLERTSGSVNSGLFYQYKYLIEGYVRDIKQYKKFNLIITGGNADLILSLIEIEFIF 240  
QY 241 NIHLTVGVRILGNSIDFKFVN 262  
Db 241 NIHLTVGVRILGNSIDFKFVN 262

## RESULT 2

US-09-813-453A-55  
; Sequence 55, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 55  
; TYPE: PRT  
; LENGTH: 256  
; ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 21.2%; Score 282; DB 9; Length 256;  
Best Local Similarity 31.3%; Pred. No. 2.6e-18;  
Matches 82; Conservative 54; Mismatches 94; Indels 32; Gaps 10;  
QY 9 LIIDIGNTSIAFALFKDQNVNLFKMTNLMRLRYDEVSFFENFDF-NVN-----KVFI 62  
Db 3 LVFDVGTNNVLGIYKGLVNVWRIKTDREKTSDEYGLISNLFYDNNVNSIDDDVII 62  
QY 63 SSVVPIILNETFNKVFISFFKIKPLFI-----GFDLNYDLTFNPKYKSKDFLGLSDVFANL 116  
Db 63 SSVVPMVMSHLENCFKYKQKQPLIVGPGIKTGLNIKYD---NPKQ-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGLINSPLNFSLLDNAYLIKFPPI 176  
Db 115 VAGIEKYGAPSL-VDFGTATFCAISERGEYLGVTIAPGKISSEALFQSASKLPVEL 173  
QY 177 STPNLLERTSGSVNSGLFYQYKYLIEGYVRDIKQYKKFN-----LIITGGNADLIL 231  
Db 174 AKPGMTICKSTVSAMQSGIYGYVGLVD---KIISMKKELNCDDVKVIATGGGLAKLIA 229  
QY 232 SLIE-IEFINHLTVGVRIL 252  
Db 230 SETKSIDYVDGF-LTLEGERII 250

## RESULT 3

US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45  
Query Match 20.8%; Score 276.5; DB 9; Length 262;  
Best Local Similarity 27.8%; Pred. No. 8.6e-18;  
Matches 71; Conservative 63; Mismatches 100; Indels 21; Gaps 7;  
QY 10 LIIDIGNTSIAFALFKDQNVNLFKMTNLMRLRYDE----VYSFFE-ENFDF-NVNKVFI 63  
Db 4 VLDVGTNAVGLVFEFEGELRQHWRMETDRKTEDEYGLMYKQLLEHGLSEDEYKGIIVS 63  
QY 64 SVPPIILNETFNKVFISFFKIKPLFI-----GFDLNYDLTFNPKYKSKDFLGLSDVFANL 117  
Db 64 SVPVPMFALERMCEYKFKIKPLVVGPGIKTGLNIKYE---NPRE-----VGADRIVNA 115  
QY 118 AAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGLINSPLNFSLLDNAYLIKFPPI 177  
Db 116 AGIHLYG-SPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIET 174  
QY 178 TPNNLLERTSGSVNSGLFYQYKYLIEGYVRDIKQYKKFNLIITGGNADLILSLIEIE 237  
Db 175 KPSSVVGKNTVSAMQSGIYGYVQVEGIVRMEKAEKQEPKVIATGGLAKLIEESNVI 234  
QY 238 FIFNIHLTVGVRIL 252  
Db 235 DVVDPELTGLGLYML 249

## RESULT 4

US-09-813-453A-47  
; Sequence 47, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Bacillus halodurans  
US-09-813-453A-47

Query Match 20.7%; Score 275.5; DB 9; Length 254;  
Best Local Similarity 28.1%; Pred. No. 1e-17;  
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;  
QY 9 LIIDIGNTSIAFALFKDQNVNLFKMTNLMRLRYDEVSFFENFDF-----NVNKVFI 62  
Db 3 LVIDVGTNTVLGYQDETLLVHHWRLATRSQKTEDEYAMTVRSFLDHAGLQFQDIDGIVI 62  
QY 63 SSVVPIILNETFNKVFISFFKIKPLFI-----GFDLNYDLTFNPKYKSKDFLGLSDVFANL 116  
Db 63 SSVVPMFMSLEQCKYFHTPMITGPGIKTGLNIKYD---NPKQ-----VGADRIVNA 114  
QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGIILGGLINSPLNFSLLDNAYLIKFPPI 176  
Db 115 VAAIIEGYVP-AIVDFGTATTTCYINEEKHYMGVITPGIMISTEALYHRSKLPRIET 173  
QY 177 STPNLLERTSGSVNSGLFYQYKYLIEGYVRDIKQYKKFNLIITGGNADLILSLIEIE 236

```
Db 174 AKPKQVGTNRIDMSQIGFYGVSDGVVKKRKAQAESEPKVIATGGLAKLIGTSEST 233
QY 237 EFIFNIHLTVGVGRIL 252
Db 234 IDVIDSFLLTLAGLQLI 249

RESULT 5
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2
```

```
Query Match 20.4%; Score 271.5; DB 9; Length 258;
Best Local Similarity 27.3%; Pred. No. 2.4e-17;
Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;

QY 9 LIIDIGNTSIAFALFKDNOVNLFIKMTNLMRLRYDEVSYFFEEFNDFN-----VNKVFPI 62
Db 3 LVIDVGNNTVGLVYHDKLELVHWRISRKTEDEFQGMILRSFDHSLMFEQIDGIII 62

QY 63 SSVVPILNETFKNVIFSFKKIKPLFI-----GFDLNYDLTFNPKYKSKFLGSDVFNAL 116
Db 63 SSVVPPINFALERCTKYFHIEPQIVGPGMTGLNKKYD---NPKE-----VGADRIYNA 114

QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPI 176
Db 115 VAAHLVG-NPLIIVDFGTATTCYIDENKQYMGGAIAPIITISTEALYSRAAKLPRIEI 173

QY 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEI 236
Db 174 TRPDNIIGKNTVSAMQSGILFYGVQVGVGVKRMKQAKODLKVIATGGLAPLIANESDC 233

QY 237 EFIFNIHLTVGVGRIL 252
Db 234 IDVIDPFLTLKGLLEI 249
```

```
RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
```

```
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 20.0%; Score 266.5; DB 9; Length 255;
Best Local Similarity 29.3%; Pred. No. 6.9e-17;
Matches 75; Conservative 48; Mismatches 112; Indels 21; Gaps 5;

QY 9 LIIDIGNTSIAFALFKDNOVNLFIKMTNLMRLRYDEVSYFFEEFNDFN-----VNKVFPI 62
Db 3 LVIDVGNNTVGLVYHDKLELVHWRISRKTEDEFQGMILRSFDHSLMFEQIDGIII 62

QY 63 SSVVPILNETFKNVIFSFKKIKPLFI-----GFDLNYDLTFNPKYKSKFLGSDVFNAL 116
Db 63 SSVVPPILTCGLERLSLGYFGMRPLVVGPGIKTGMPLOYD---NPKE-----VGADRIYNA 114

QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPI 176
Db 115 VAGYEKIR-TSLIIVDFGTATTCYIDENKQYMGGAIAPIITISTEALYSRAAKLPRIEI 173

QY 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADLILSLIEI 236
Db 174 IRPSAIIRANTVNSMQAGIYGVGLVDEIVTRKMAESKDAPRVIATGGLASLIAPESKT 233

QY 237 EFIFNIHLTVGVGRIL 252
Db 234 IEAVEEYLTLEGLRIL 249
```

```
RESULT 7
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3
```

```
Query Match 19.7%; Score 262; DB 9; Length 250;
Best Local Similarity 30.1%; Pred. No. 1.8e-16;
Matches 72; Conservative 54; Mismatches 87; Indels 26; Gaps 9;

QY 9 LIIDIGNTSIAFALFKDNOVNLFIKMTNLMRLRYDE---VYSFFEEFNDFN-VNKVFPI 62
Db 20 LVIDVGNNTVGLVYHDKLELVHWRISRKTEDEFQGMILRSFDHSLMFEQIDGIII 79

QY 63 SSVVPILNETFKNVIFSFKKIKPLFI-----GFDLNYDLTFNPKYKSKFLGSDVFNAL 116
Db 80 SSVVPMYSLHEMINKYFKINPLVVGPGIKTGINKYD---NPKE-----VGADRIYNA 131

QY 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPI 176
Db 132 VAAHEIYK-RSLIIVDFGTATTCYIDENKQYMGGAIAPIITISTEALYSRAAKLPRIEI 190
```





## RESULT 11

US-09-813-453A-6  
 ; Sequence 6, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 258  
 ; TYPE: PRT  
 ; ORGANISM: Rhodobacter capsulatus  
 US-09-813-453A-6

Query Match 14.8%; Score 197.5; DB 9; Length 258;  
 Best Local Similarity 23.7%; Pred. No. 1.6e-10;  
 Matches 60; Conservative 55; Mismatches 123; Indels 15; Gaps 6;  
 QY 9 LIIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFPE-----ENDFNVNKFVIS 63  
 DB 3 LCIDCGNTVFSVMDGTDFAATWRIATDHRRTADEYVVLNLMQLKLGGRISALIS 62  
 QY 64 SVVPILNETKNIFFSFKIKPLFIQFDLNYDLTFNFYKSKDKFLGSGDVFANLVAATENY 123  
 DB 63 STAPRVVNLNLCNRYFDCRPYVG-KPGCELPVAPRVDPGTTVGPDLVNTVAGYDRH 121  
 QY 124 SFENVLVVDGTACTIFAVSRQDGLGIINSGLINFSPLNLSLDNAYLIKPKPISTPNLL 183  
 DB 122 G-GDLIVVDGTATTFDVPVAPDGAIGVIAAGVNLGSLALHMAAALPHVDVTKPGQVI 180  
 QY 184 ERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEF-IFN- 241  
 DB 181 GTNIVACIQSGVWGYIGLVGIVRQIRMERDRPMKVIATGG-----LASLFDLGFDFDK 236  
 QY 242 --IHLVVEGVRIIL 252  
 DB 237 VEDDLTMHGLRLI 249

## RESULT 12

US-09-813-453A-57  
 ; Sequence 57, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 57  
 ; LENGTH: 219  
 ; TYPE: PRT  
 ; ORGANISM: Dehalococcoides ethenogenes

## US-09-813-453A-57

Query Match 14.3%; Score 190; DB 9; Length 219;  
 Best Local Similarity 25.0%; Pred. No. 6.4e-10;  
 Matches 58; Conservative 45; Mismatches 87; Indels 42; Gaps 6;  
 QY 1 MNKPLLSLIIDIGNTSIAFALFK-----DNQVNLFIKMKTNLMRLRY 42  
 DB 1 MSELKVA---VDIGNTSVNIIGIFEKGLLANHNLGSLVQAORMADEVASLGLLQHGAIHP 57  
 QY 43 DEVYSFEEENFDNVNKFVISVVPILNETKNIFFSFKIKPLFIQFDLNYDLTF---N 99  
 DB 58 EE-----LNRVIMCSVVPPLTTTTEEVEKSYFKAAPLVVVGAIKSGVKYRMDN 105  
 QY 100 PYKSKDKFLGSGDVFANLVAATENYSFENVLVVDLTACTIFAVSRQDGLGIINSGLPI 159  
 DB 106 PRE-----VGADRIVNAARVLYP-GACIIIDMGTTATFTDLSEGGAYIGGAIPGIAT 159  
 QY 160 NFNSLLDNAYLIKPKPISTPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIK 211  
 DB 160 SAQAIAETSKLPKIEIIRPAKVIQSGNTVSAMQSGIYFYGIVGLVEELVRRIO 211

## RESULT 13

US-09-813-453A-51  
 ; Sequence 51, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT APPLICATION NUMBER: US/09/813,453A  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: US 60/227,860  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: 09/667,569  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 51  
 ; LENGTH: 260  
 ; TYPE: PRT  
 ; ORGANISM: Caulobacter crescentus  
 US-09-813-453A-51

Query Match 14.2%; Score 189.5; DB 9; Length 260;  
 Best Local Similarity 26.9%; Pred. No. 9e-10;  
 Matches 67; Conservative 41; Mismatches 130; Indels 11; Gaps 4;  
 QY 9 LIIDIGNTSIAFALFKDQVNLFIKMKTNLMRLRYDEVYSFEEENFDN-----VNKVF 62  
 DB 3 LAIEQGNITMTFAIHGASWVAQWSSATESRTADEYVVLWSQLLSMQGLGFRADAVII 62  
 QY 63 SVVPILNETKNIFFSFKIKPLFIQFDLNYDLTFNPKYKSKDKFLGSGDVFANLVAIEN 122  
 DB 63 SVVPQSIIFNLNLSRRYFNVEPLVIGENAKLIDVRIEKPSE--AGADRLVNAIGAAMV 120  
 QY 123 YSFENVLVVDGTACTIFAVSRQDGLGIINSGLINFSPLNLSLDNAYLIKPKPISTP--N 180  
 DB 121 YP-GPLVVIDSGTATTFDVAADGAFEGGIITAPGINLSMQALHMAAALPRIAIOIPAGN 179  
 QY 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEF 240  
 DB 180 RIVGTDVSAQSGVFWGYISLIEGLVARIAERPEPTVINTATGGVASFEGATSDIDH 239  
 QY 241 NIHLTVVEGV 249  
 DB 240 DSDLTIRGL 248

## RESULT 14

US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 265  
; TYPE: PRN  
; ORGANISM: Streptomyces coelicolor  
; S-09-813-453A-4

Query Match 13.6%; Score 181.5; DB 9; Length 265;  
Best Local Similarity 23.6%; Pred. No. 5e-09;  
Matches 63; Conservative 55; Mismatches 114; Indels 35; Gaps 7;  
QY 9 LIIDIGNTSIAFALFKDQNVNLFIRKMTNMLRYDE-----VYSFFEFNFDFNVN 58  
DB 3 LIIDVGNTHVLGLFDGDIHWRISTDSRRTADELAVLLOGLMGHPLGLDGLGDID 62  
QY 59 KVFISVVPILNETFKNVIFSFKKIKPLFI-----GFDLNDLTFNPKYKSKFLGSD 111  
DB 63 GIAICATVPSVLHRELRETRRYGDPVAVLPEPGVKTGVP-----LTDHPKE-----VGAD 114  
QY 112 VFANLVAALENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLINFLNSLLDNAYLI 171  
DB 115 RIINAAVELYG-GPAIVVDFGTATFDVAVSARGEYIGGVIAPIEISVEALGVKGAQL 173  
QY 172 KKFPISTPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQMYK-----KFNLIITGG 225  
DB 174 RKEIVARPSVICKNTVEAMQSGIVYGFAGQVDGV---VNRMARELADDDVTVIATGG 230  
QY 226 NADLILSLIEFIFNHLTVGVRIL 252  
DB 231 LAPMVLGESSVIDEHEPWLTLMLRLV 257

RESULT 15  
3-09-813-453A-10  
; Sequence 10, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 273  
; TYPE: PRN  
; ORGANISM: Treponema pallidum  
US-09-813-453A-10

Query Match 12.8%; Score 170; DB 9; Length 273;

Best Local Similarity 25.3%; Pred. No. 6e-08;  
Matches 64; Conservative 50; Mismatches 123; Indels 16; Gaps 6;  
QY 9 LIIDIGNTSIAFALFKDQNVNLFIRKMTNMLRYDEYISFFEFNF-----DFNV 57  
DB 2 LIIDVGNTHVLGLFDGDIHWRISTDSRRTADELAVLLOGLMGHPLGLDGLGDID 60  
QY 58 NKVFISVVPILNETFKNVIFSFKKIKPLFI-----GFDLNDLTFNPKYKSKFLGSDVFNALV 117  
DB 61 RDAFISVVPILNETFKNVIFSFKKIKPLFI-----GFDLNDLTFNPKYKSKFLGSDVFNALV 120  
QY 118 AALENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLINFLNSLLDNAYLIKPKPIS 177  
DB 121 AAYVHER-SACVVVDCGTALTFTAVDGTGLIQGVIAPIEISVEALGVKGAQL 179  
QY 178 TPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQMYKFNLIITGGNADLILSLIE 237  
DB 180 LPDSVLGKDTTHAVQAGVVRGTLEFVIRAMIAQCOKELGCRCAAVITGGLSLFSS--EVD 237  
QY 238 F-IFNHLTVGVRIL 249  
DB 238 FPDIDQLTSLGL 250

Search completed: June 24, 2003, 22:23:59  
Job time: 12.9453 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.3688 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSLLIDIGNTSIAF.....HLTVEGVRIILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- SPTREMBL\_21.\*
- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	100.0	262	16	O51477 borrelia bu
2	293.5	22.1	255	16	Q8r7m2 thermoaer
3	289.5	21.8	273	16	Q97eb4 clostridium
4	275.5	20.7	254	16	Q9kb5 bacillus ha
5	253.5	19.0	259	16	Q92f54 listeria in
6	248.5	18.7	259	16	Q8yac5 listeria mo
7	248.5	18.7	259	16	Q8xhl5 clostridium
8	242.5	18.2	258	2	Q9f985 bacillus st
9	232.5	17.5	256	16	Q8rfe4 fusobacteri
10	203	15.3	246	16	Q9wzy5 thermotoga
11	189.5	14.2	261	16	Q9a6z1 caulobacter
12	181.5	13.6	265	16	Q9x8n6 streptomyce
13	170	12.8	273	16	Q83446 streptococ
14	159	11.9	262	16	Q9rx54 deinococcus
15	152.5	11.5	224	16	Q98q93 mycoplasma
16	143	10.7	274	16	Q9cd56 mycobacteri

17	141.5	10.6	229	16	O67753
18	138	10.4	272	16	O06282
19	121.5	9.1	209	16	Q9pia9
20	121.5	9.1	476	8	Q35809
21	119	8.9	517	16	Q8xmy0
22	118	8.9	276	16	O8yqd7
23	118	8.9	625	10	Q98s73
24	118	8.9	765	8	Q9MTD7
25	116.5	8.8	553	17	Q975l9
26	116	8.7	2178	2	Q46149
27	114	8.6	815	12	Q9EN43
28	112.5	8.5	963	16	Q98r23
29	112	8.4	636	10	Q98RR1
30	111.5	8.4	629	12	Q9YW36
31	111	8.3	212	2	Q32514
32	110.5	8.3	270	2	Q9AIY7
33	110.5	8.3	1003	16	Q9PQ01
34	110	8.3	654	12	Q9EN03
35	103.5	8.2	1179	5	O8T2A2
36	107.5	8.1	636	2	Q9AHW5
37	107.5	8.1	1613	10	Q98S55
38	107	8.0	477	5	O97280
39	107	8.0	590	10	Q98S67
40	107	8.0	842	2	Q93U45
41	104.5	7.9	267	2	Q9AI23
42	104.5	7.9	447	12	Q91ML6
43	104.5	7.9	448	12	Q9EMG8
44	104	7.8	346	16	O8RHE3
45	104	7.8	660	16	O8RGT2

ALIGNMENTS

RESULT 1

O51477	PRELIMINARY;	PRT;	262 AA.
AC	O51477;		
DT	01-JUN-1998 (TREMBlrel. 06, Created)		
DT	01-JUN-1998 (TREMBlrel. 06, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	Hypothetical protein BB0527.		
GN	BB0527		
OS	Borrelia burgdorferi (Lyme disease spirochete).		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.		
OX	NCBI_TaxID=139;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 35210 / B31;		
RX	MEDLINE=98065943; PubMed=9403685;		
RA	Fraser C.N., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,		
RA	Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,		
RA	Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,		
RA	Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,		
RA	van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,		
RA	Utterback T., Watney L., McDonald L., Artiach P., Bowman C.,		
RA	Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,		
RA	Smith H.O., Venter J.C.;		
RT	"Genomic sequence of a Lyme disease spirochaete, Borrelia		
RT	burgdorferi."		
RL	Nature 390:580-586(1997).		
DR	EMBL; AE001154; AAC66882.1; -		
DR	TIGR; BB0527; -		
DR	InterPro; IPR004619; Baf.		
DR	Pfam; PF03309; Bvg_acc_factor; 1.		
DR	TIGRfams; TIGR00671; baf; 1.		
KW	Hypothetical protein; Complete proteome.		
SQ	SEQUENCE 262 AA; 29822 MW; DC8CD558C16B9AF CRC64;		

Query Match 100.0%; Score 1331; DB 16; Length 262;  
Best Local Similarity 100.0%; Pred. No. 2.9e-90;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MNPPLSELLIDIGNTSIAFALPKDNOVNLFIKMTNLMRLRYDEVYSFFENPDFNVNKV 60
Db 1 MNPPLSELLIDIGNTSIAFALPKDNOVNLFIKMTNLMRLRYDEVYSFFENPDFNVNKV 60
QY 61 FISSVVPILNETKFNIFSFYKIPKPLFIFGDLNLDLTFFNPKSKDKFLGSDVFNVAAL 120
Db 61 FISSVVPILNETKFNIFSFYKIPKPLFIFGDLNLDLTFFNPKSKDKFLGSDVFNVAAL 120
QY 121 ENTSFENVLVVDIGTACTIPAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPISTPN 180
Db 121 ENTSFENVLVVDIGTACTIPAVSRQDGLGILGILGINSGLPLNFNSLLDNAYLIKFPPISTPN 180
QY 181 NLERETSGSVNSGLFYQYKYLLEGVYRDIKOMYKKFNLIITGGNADLILSLIEFIF 240
Db 181 NLERETSGSVNSGLFYQYKYLLEGVYRDIKOMYKKFNLIITGGNADLILSLIEFIF 240
QY 241 NIHLTVEGVRLGNSIDFKFVN 262
Db 241 NIHLTVEGVRLGNSIDFKFVN 262
RESULT 2
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WB4T / JCM11007;
RX MEDLINE-21992816; PubMed-11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013180; AA025520.1; -
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CAGED CRC64;
Query Match 22.1%; Score 293.5; DB 16; Length 255;
Best Local Similarity 29.7%; Pred. No. 4.4e-14;
Matches 76; Conservative 53; Mismatches 106; Indels 21; Gaps 5;
QY 9 LIIDIGNTSIAFALPKDNOVNLFIKMTNLMRLRYDEVYSFFENPDFN-----VNVKVI 62
Db 3 LAFDVGNTNIVGVFKKLLHSFRISTDKNTYDEYGLMVLNOLGYNGLSITIEDVII 62
QY 63 SSVVPILNETKFNIFSFYKIPKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFNAL 116
Db 63 SSVVPILNETKFNIFSFYKIPKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFNAL 116
QY 117 VAAIENYSPENVLVVDIGTACTIPAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPI 176
Db 115 VAAIENYSPENVLVVDIGTACTIPAVSRQDGLGILGILGINSGLPLNFNSLLDNAYLIKFPPI 176
QY 177 STPNLLERTTSGSVNSGLFYQYKYLLEGVYRDIKOMYKKFNLIITGGNADLILSLIEI 236
Db 174 TTPPTVINNTVASMOSGLIYGVGNVYIVTRMKGEFAPSAYVATGFGFANNIAESKT 233
QY 237 EFIFNIHLTVEGVRL 252
Db 234 IDTVNEMLTLEGLRII 249
```

```
RESULT 3
Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
GN CAC3200.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE-21359325; PubMed-11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007815; AAK81136.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; Baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;
Query Match 21.8%; Score 289.5; DB 16; Length 273;
Best Local Similarity 29.6%; Pred. No. 9.5e-14;
Matches 77; Conservative 62; Mismatches 96; Indels 25; Gaps 8;
QY 9 LIIDIGNTSIAFALPKDNOVNLFIKMTNLMRLRYDE----VYSFEEF-EDFN-VNVKVI 62
Db 14 LVLVDVGNTNIVGLIYNDTKLTAEWRLSTDVLSADEYGIQVWNLFOQDKLDTLVEGVII 73
QY 63 SSVVPILNETKFNIFSFYKIPKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFNAL 116
Db 74 SSVVPILNETKFNIFSFYKIPKPLFI-----GFDLNYDLTFNPKSKDKFLGSDVFNAL 125
QY 117 VAAIENYSPENVLVVDIGTACTIPAVSRQDGLGILGINSGLPLNFNSLLDNAYLIKFPPI 176
Db 126 VAAIENYSPENVLVVDIGTACTIPAVSRQDGLGILGILGINSGLPLNFNSLLDNAYLIKFPPI 184
QY 177 STPNLLERTTSGSVNSGLFYQYKYLLEGVYRDIKOMYKKFNLIITGGNADLILSLIEI 232
Db 185 IKPAPAICKNTISSIGSIVGYGVYIVVERKKELOEGEKEPLVATGGLAKLISE 244
QY 233 LIEIFIFNIHLTVEGVRL 252
Db 245 EAKNVVDVNPFLTLEGLRII 264
RESULT 4
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 20.7%; Score 275.5; DB 16; Length 254;
Best Local Similarity 28.1%; Pred. No. 9.3e-13;
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;

Y 9 LIIDIGNTSIAFALEKQNVNLFKMTNMLRYDEVVSFFPEPDF-----NVNKVFI 62
   I:::||||: ::::: ::::: || || || || || || || || || || ||
3 LVIDVGNTNCTVGVYKEQKLLRHWRMTDRHRTSDELGMVTLNFFSYANLTSDIOGIII 62
   |||||: ::::: ::::: || || || || || || || || || || ||
63 SSVVPILNETFKNVIFSFFKIKPLFI-----GFDLNYDLFNPYKSKFLLGSDVFANL 116
   |||||: ::::: ::::: || || || || || || || || || || ||
63 SSVVPPMMFSLQCMCKKYFHTVPMIIGIKTGLNRYD---NPKR-----VGADRIVNA 114
   |||||: ::::: ::::: || || || || || || || || || || ||
117 VAAIENYSFENVLVDLGTACTIFAVSRDGLGILGINSGLPLNFNSLIDNAYLIKPKPI 176
   |||||: ::::: ::::: || || || || || || || || || || ||
115 VAAELVGYG-ATVVDGTATATYCLINEKKQYAGGVATPAGIMISTEALYHRASKLPRIE 173
   |||||: ::::: ::::: || || || || || || || || || || ||
177 STPNLLERTTSGSVNSGLFYQKYLIEGVYRDIKMTKKFNLIITGNGADLILSLIEI 236
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
174 AKPQVGVGTNTIDSMQGFYGVYQVSDGVVVKRMAQAASEPKVIATGLAKLIGTSEET 233
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
237 EFIFNIHLTVGVGRIL 252
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
234 IDVIDSFLTLKGLQLI 249
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||

RESULT 5
Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95486.1; -
DR ListList; LIN00253; -

Query Match 18.7%; Score 248.5; DB 16; Length 259;
Best Local Similarity 27.7%; Pred. No. 9.1e-11;
Matches 70; Conservative 53; Mismatches 115; Indels 15; Gaps 5;
```

```
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0CEFA64F CRC64;

Query Match 19.0%; Score 253.5; DB 16; Length 259;
Best Local Similarity 28.1%; Pred. No. 3.9e-11;
Matches 71; Conservative 53; Mismatches 114; Indels 15; Gaps 5;

QY 9 LIIDIGNTSIAFALEKQNVNLFKMTNMLRYDEVVSFFPE-----NVNKVFI 62
   I:::||||: ::::: ::::: || || || || || || || || || || ||
3 LVIDVGNTNCTVGVYKEQKLLRHWRMTDRHRTSDELGMVTLNFFSYANLTSDIOGIII 62
   |||||: ::::: ::::: || || || || || || || || || || ||
63 SSVVPILNETFKNVIFSFFKIKPLFI-----GFDLNYDLFNPYKSKFLLGSDVFANL 119
   |||||: ::::: ::::: || || || || || || || || || || ||
63 SSVVPPIMHAMETMCRVFNIRPLIVGPGIKTGLNKLVDNPRE-----IGSDRIVNAVA 117
   |||||: ::::: ::::: || || || || || || || || || || ||
120 IENYSFENVLVDLGTACTIFAVSRDGLGILGINSGLPLNFNSLIDNAYLIKPKPISTP 179
   |||||: ::::: ::::: || || || || || || || || || || ||
118 SEEYG-TPVIVVDFGTATTCYIDEAGVYQGAIAFGIMISTEALYNRAAKLPRVDIAES 176
   |||||: ::::: ::::: || || || || || || || || || || ||
180 NLLERTTSGSVNSGLFYQKYLIEGVYRDIKMTKKFNLIITGNGADLILSLIEIFI 239
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
177 SOIISKSTVASMAGIFGIFGICQCGEIIAEMKQSNTPVWVATGGIARMIKTESSAVDI 236
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
240 FNIHLTVGVGRIL 252
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||
237 LDPFTLKGLELL 249
   : : : : : |||||: ::::: ::::: || || || || || || || || || ||

RESULT 6
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1; -
DR MEROPS; M41.009; -
DR ListList; LMO00221; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 18.7%; Score 248.5; DB 16; Length 259;
Best Local Similarity 27.7%; Pred. No. 9.1e-11;
Matches 70; Conservative 53; Mismatches 115; Indels 15; Gaps 5;
```

```

QY 9 LIIDIGNTSIAPALFKDQNVNFIKMTNMLRYDE----VVSFFE--ENFDNVNKVFI 62
Db 3 LVIDGNTCTGVTEKQKLLKHWRTTDRHRTSDLGMTVLNFFSYANLTPSDIQGIII 62
QY 63 SSVVPILNETFNKVFISFFKIKPLFIIGFDLNDLTF---NPYKSKDFLIGSDVFANLVA 119
Db 63 SSVVPIMHAMETCMVRYFNIRPLIVPGIKTGLNKLKVDNPRE-----IGSDRIVNAVAA 117
QY 120 IENYSFENVVLDLGTACTIFAVSRDQDILGILGINSGLINFLNSLNDNAYLIKPFISIP 179
Db 118 SEEGY-TPVIVDFGTATTCVDESQVVGAGIAPGIMISYEALYNRAKIPRVDIAS 176
QY 180 NLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIFI 239
Db 177 SQIICKTSVSSMQAGIFYGVQCEGIIAEMKKQSNASPVVATGGLARMITEKSSAVDI 236
QY 240 FNIHLTVGVRIL 252
Db 237 LDPFLTLKLELL 249

RESULT 7
Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Hypothetical protein CP2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shmizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -
DR InterPro; IPR000515; BPD.transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
W PROSITE; PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
W Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

```

```

Query Match 18.7%; Score 248.5; DB 16; Length 259;
Best Local Similarity 26.4%; Pred. No. 9.1e-11;
Matches 69; Conservative 69; Mismatches 96; Indels 27; Gaps 8;

QY 9 LIIDIGNTSIAPALFKDQNVNFIKMTNMLRYDE----VVSFFE--ENFDNVNKVFI 61
Db 3 LVIDGNTCTGVTEKQKLLKHWRTTDRHRTSDLGMTVLNFFSYANLTPSDIQGIII 61
QY 62 ISSVVPILNETFNKVFISFFKIKPLFI-----GFDLNDLTFNPKSKDFLIGSDVFAN 115
Db 62 ISSVVPIMHSLNENVRCKEPIVVGPIKTIKINIKYD---NPKE-----VGADRIYN 113
QY 116 LVAATENYFENVVLDLGTACTIFAVSRDQDILGILGINSGLINFLNSLNDNAYLIKPF 175
Db 114 AVAAPEKHK-KPMIITDEGTATTCFAITEKDGILGNGICPGQISADALFERAAKLPRIE 172
QY 176 ISTPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLIL 231
Db 173 LEKPKSVICKNTVSMQAGIIFYGVQCEGIIIVKRMKEMMDIGERPEFVLATGGLAKLY 232

```

```

QY 232 SLIEIEFIFNIHLTVGVRIL 252.
Db 233 SETDVIDEVRDKLTLEGLKIL 253

RESULT 8
Q9F985 PRELIMINARY; PRT; 258 AA.
AC Q9F985;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RA Vasquez C., Pichuanes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198621; AAG28531.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 18.2%; Score 242.5; DB 20; Length 258;
Best Local Similarity 25.3%; Pred. No. 2.5e-10;
Matches 65; Conservative 60; Mismatches 109; Indels 21; Gaps 5;

QY 10 IIDIGNTSIAPALFKDQNVNFIKMTNMLRYDEVYFEEFDF-----NVNKFVIS 63
Db 4 VLDVGNNTVGVYDGDGLKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGIIS 63
QY 64 SVPILNETFNKVFISFFKIKPLFI-----GFDLNDLTFNPKSKDFLIGSDVFANLY 117
Db 64 SVPPIPMFALERMCIFYFKIPIIVGPIKTIKINIKYD---NPKE-----VGADRIYNV 115
QY 118 AAIENYSFENVVLDLGTACTIFAVSRDQDILGILGINSGLINFLNSLNDNAYLIKPFIS 177
Db 116 AGIHLVG-SPLIIVDFGTATTCYIINEHKQYMGAGIAPGIMISTEALFARAALPRIEIA 174
QY 178 TPNNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIE 237
Db 175 RPDDIIGKNTVSMQAGIIFYGVQCEGIVSRMKAKSPVPKVIATGGLASLASESNVI 234
QY 238 FIFNIHLTVGVRIL 252
Db 235 DIVDPELTTLGLKIL 249

RESULT 9
Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,

```

```

RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RL nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
KW EMBL; AE010586; AAL94957.1; -
KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 17.58; Score 232.5; DB 16; Length 256;
Best Local Similarity 25.6%; Pred. No. 1.3e-09;
Matches 64; Conservative 65; Mismatches 110; Indels 11; Gaps 5;

QY 11 IDIGNTSIAFALFKDN-QVNLFKMKTNLMRLRYDEVYSEFFENFDN-----VKNYFIS 63
DB 5 IDIGNTHIVGIYDNNGELSTFRIATNDKMTDEYEFYFNITKYNEISIKKYDAILIS 64
QY 64 SVVPILNETPKNVIFSEFKIKPLFIQFDNLVDFNFYKSKDFL-LGSDVFANLVAIAEN 122
DB 65 SVVPILNETPKNVIFSEFKIKPLFIQFDNLVDFNFYKSKDFL-LGSDVFANLVAIAEN 122
QY 123 YSFENVLVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKKEPISTPNLLR 182
DB 123 YSFENVLVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKKEPISTPNLLR 182
QY 123 YPKNLVIFDGTG-TTYDVLKGVYIGGGILPGIDMSINALYGNATKLPKRVKFTTSSV 181
DB 123 YPKNLVIFDGTG-TTYDVLKGVYIGGGILPGIDMSINALYGNATKLPKRVKFTTSSV 181
QY 183 LERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIFNI 242
DB 183 LERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIFNI 242
QY 182 LGTDTMKQIQALFFGAGQIKHIIKINEELNEIFVLATFGGLGKILSAEIDEYDA 241
DB 182 LGTDTMKQIQALFFGAGQIKHIIKINEELNEIFVLATFGGLGKILSAEIDEYDA 241
QY 243 HLATVEGVRIIL 252
DB 243 HLATVEGVRIIL 252
QY 242 NLSLKGLTYL 251
DB 242 NLSLKGLTYL 251

RESULT 10
Q9WZY5
ID Q9WZY5 PRELIMINARY; PRT; 246 AA.
AC Q9WZY5;
DT 01-NOV-1999 (TREMREL. 12, Created)
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -
DR TIGR; TM0883; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 15.38; Score 203; DB 16; Length 246;
Best Local Similarity 24.6%; Pred. No. 1.9e-10;
Matches 60; Conservative 62; Mismatches 112; Indels 10; Gaps 7;

QY 9 LIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSEFFENFDN-----VKNVFI 65
DB 3 LLDVGNTHSVFSITEDGKTRRWRLSTGVQTEDELFLHLLGLDAMREIKGIGVASV 62

QY 66 VPILNETPKNVIFSEFKIKPLFIQFDNLVDFNFYKSKDFL-LGSDVFANLVAIAENYSEF 125
DB 63 VPTONTVIERFSQYKHISPIWVAK-NGCVKNNKNPSE--VGADRVANVAVFYKEYG- 118
QY 126 ENVLVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKKEPISTPNLLER 185
DB 119 KNGIIMGTATTVDLVV-NGSEGAAILPGFMMVHSLFRCTAKLPLVEVPADPVGK 177
QY 186 TTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIFNIHLT 245
DB 178 DTEENIRLGVVNSGYALEGIIGRIKEVY-GDLPVVLTTGGQSKVKMDIKHE-IFDEDLT 235
QY 246 VEGV 249
DB 236 IKGV 239

RESULT 11
Q9A6Z1
ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.
AC Q9A6Z1;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -
DR TIGR; CC1935; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 14.2%; Score 189.5; DB 16; Length 261;
Best Local Similarity 26.9%; Pred. No. 2e-06;
Matches 67; Conservative 41; Mismatches 130; Indels 11; Gaps 4;

QY 9 LIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYSEFFENFDN-----VKNVFI 62
DB 4 LAIEQGNNTMFAIHGASWVAQWRSATESRTADEYVVMVLSQLLSMOGLGFRDAIVII 63
QY 63 SSVVPILNETPKNVIFSEFKIKPLFIQFDNLVDFNFYKSKDFL-LGSDVFANLVAIAEN 122
DB 64 SSVVPQSTFNLRLNLSRRYFNVEPLVIGENAKIGDVRIEKPSE--AGADRLVNAIGAMV 121
QY 123 YSFENVLVDLGTACTIFAVSRQDGIILGGINSGPLINFNSLLDNAYLIKKEPISTPN 180
DB 122 YP-GLVVIDSGTATTFDVAADGAFEGGIAPGINLSQALHEAAKLPRIAIQRPAGN 180
QY 181 NLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIF 240
DB 181 RIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVASLFEAGTSDIDHF 240
QY 241 NIHLTVEGV 249

```

## RESULT 14

Q9RX54  
ID 09RX5



```
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rl;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum C.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
A "Genome sequence of the radioresistant bacterium Deinococcus
T radiodurans R1.";
RT Science 286:1571-1577(1999).
RL EMBL; AE001905; AAF10040.1; -.
DR DR0461.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 11.9%; Score 159; DB 16; Length 262;
Best Local Similarity 23.2%; Pred. No. 0.00035;
Matches 59; Conservative 51; Mismatches 132; Indels 12; Gaps 7;

QY 4 PLISELIIDIGNTSIAFALEF-KNQVNLFFKMTNLMRLRYDEVYSFEENFDNFN----VN 58
DB 2 PAPPLAVDIGNTTVTLGLDASGALHTWRIETNREMLPDDLALQLHGLFTLAGAPIR 61
QY 59 KVFISVPIINTEFNKVFISFFKIKPLFGDLNLTENPYKSKDKFLGSLGVFNALVA 118
DB 62 AAVLSSVAPPVGENYALAKRHMIDAFVSAENLPDVTVE--LDTPGSGADRCLNLF 119
QY 119 AIENY--SFENVLVDLGTACTIFAVSRQDGLGGIINSGLINFNLSLLDNAYLIKKEPI 176
DB 120 A-EKYLGLDYAVVDFGTSTNFDVVGRRRFLGGILATGAQVVSADALFARAACLPRITL 178
QY 177 STPNLLERTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGGNADILSLI-E 235
DB 179 QAPETAIGNTVHALQSLGVFGYAEVMDGLLRIRAEPLGCAVAVATGGFSRTVQICQE 238
QY 236 IEFENHLTVGEV 249
DB 239 IDY-IDETLTLRGL 251

RESULT 15
Q98093 PRELIMINARY; PRT; 224 AA.
AC Q98093;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE Hypothetical protein MYP0_4730.
GN MYP0_4730.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RL Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13646.1; -.
DR MYP0_4730; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 224 AA; 25363 MW; 240600CED6AB2B4B CRC64;

Query Match 11.5%; Score 152.5; DB 16; Length 224;
Best Local Similarity 26.2%; Pred. No. 0.00087;
Matches 62; Conservative 44; Mismatches 90; Indels 41; Gaps 12;

QY 8 ELIIDIGNTSTAPALFADKDNQVNLFFKMTNLMRLRYDEVYSFEENFDNFN---VNVKVFIS 63
DB 2 KIFIDLGNFPAKFAKKNENKAHFLRLKTSQVSDPSFEIKSF--NLFEFNKLDVKEILIC 59
QY 64 SVVPILNETFNKVFISFFKIKPLFGDLNLTENPYKSKDKFL-----LGSDVF 113
DB 60 SV-----RNAKENQILE-FKLKSI-----KNAKIDFFIHKSLVKFCOKELTSEIGLDIV 110
QY 114 ANLVAALENYSFENVLVDLGTACTIFAVSRQD-----GILGGIINS-GPLINFNLSLLNA 168
DB 111 ANAYVYL--HKSNAIFISLGTAVITQIKNSSIEGTSIYFGIYOSKFNFFNVVAKIESN 168
QY 169 YLIKFFPISTPNLLERTSGVNSGLFYQYKYLIEGVYRDIKQYKKKFNLIITGG 225
DB 169 FNIIKIP-----PILGKTTLESISLVRGVSFLKGVIDEI----DKTSDIFITGG 216

Search completed: June 24, 2003, 21:59:27
Job time : 23.3688 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.65383 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPILLSLIDIGNTSIAF.....HLTVEGVRIILGNSIDPKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---Searched: 262574 seqs, 29422922 residues

,otal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	121.5	9.1	476	4	US-09-316-083-3
2	99.5	7.5	410	4	US-09-140-466-2
3	96.5	7.3	3200	2	US-08-477-451-8
4	95	7.1	493	4	US-08-686-988C-4
5	92.5	6.9	284	4	US-09-134-001C-2927
6	92	6.9	331	4	US-09-134-001C-3626
7	91.5	6.9	574	4	US-09-134-001C-5409
8	89	6.7	3079	5	PCT-US94-00198-4
9	87	6.5	1854	4	US-09-004-838-108
10	86	6.5	275	2	US-08-645-193B-19
11	86	6.5	299	4	US-09-134-001C-3433
12	84.5	6.3	425	4	US-09-134-001C-5654
13	84.5	6.3	1095	4	US-09-206-942-69
14	84.5	6.3	1536	1	US-08-038-682-2
15	84.5	6.3	1536	1	US-08-302-832-2
16	84.5	6.3	1536	2	US-08-530-198-2
17	84.5	6.3	1536	2	US-08-469-880-2
18	84.5	6.3	1536	2	US-08-728-470-2
19	84.5	6.3	1536	2	US-08-617-697-2
20	84.5	6.3	1536	4	US-08-719-641-2
21	84.5	6.3	1536	4	US-09-206-942-67
22	83	6.2	607	4	US-09-537-682-1
23	83	6.2	873	3	US-09-187-331-6
24	83	6.2	873	4	US-09-470-946-6
25	83	6.2	873	4	US-09-438-906-6
26	83	6.2	873	4	US-09-438-906-4
27	83	6.2	925	2	US-08-392-946-1

28	83	6.2	925	2	US-08-504-169-1	Sequence 1, Appl1
29	83	6.2	925	5	PCT-US94-14893-1	Sequence 1, Appl1
30	82.5	6.2	293	1	US-08-118-270-60	Sequence 60, Appl1
31	82.5	6.2	293	5	PCT-US93-08528-60	Sequence 60, Appl1
32	82.5	6.2	694	2	US-08-895-522-4	Sequence 4, Appl1
33	82.5	6.2	694	3	US-09-195-391-4	Sequence 4, Appl1
34	82	6.2	194	3	US-08-741-411-9	Sequence 9, Appl1
35	82	6.2	218	2	US-08-531-525-19	Sequence 19, Appl1
36	82	6.2	218	2	US-08-718-270A-19	Sequence 19, Appl1
37	82	6.2	451	4	US-09-371-913A-9	Sequence 9, Appl1
38	82	6.2	1180	4	US-09-206-942-63	Sequence 63, Appl1
39	82	6.2	1188	4	US-09-206-942-63	Sequence 4378, Ap
40	81	6.1	878	4	US-09-134-001C-4378	Sequence 5, Appl1
41	80.5	6.0	717	4	US-08-924-629C-5	Sequence 53, Appl1
42	80.5	6.0	977	4	US-09-206-942-53	Sequence 51, Appl1
43	80.5	6.0	983	4	US-09-206-942-51	Patent No. 5352450
44	80.5	6.0	1565	6	5352450-2	Sequence 11, Appl1
45	79.5	6.0	220	1	US-07-991-867B-11	

ALIGNMENTS

RESULT 1  
US-09-316-083-3  
; Sequence 3, Application US/09316083A  
; Patent No. 6280942  
; GENERAL INFORMATION:  
; APPLICANT: The Institute of Physical and Chemical Research  
; TITLE OF INVENTION: Endonuclease  
; FILE REFERENCE: PH-651  
; CURRENT APPLICATION NUMBER: US/09/316,083A  
; CURRENT FILING DATE: 1999-05-20  
; EARLIER APPLICATION NUMBER: JP98/141861  
; EARLIER FILING DATE: 1998-05-22  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-316-083-3

Query Match	9.1%	Score 121.5	DB 4	Length 476
Best Local Similarity	20.2%	Pred. No. 8.8e-05		
Matches	73	Conservative 55	Mismatches 99	Indels 135
Gaps	19			
QY	15	NTSIAFALFKONQVNLTKMTNLMRYDE	-----VYSFF	49
DB	46	NMTCFIKWNNKILLDDMYNLYNHKQTPMSKRLMSKNIMDYKLLTYFVLN	105	
QY	50	-----EENFDFNVNKFVSSVPIINETKKNVIFSFYKPKPLFGFDLNY	-----DLT	97
DB	106	KMKEMDYNNNNNNSLSK-----YNELKNIM	-----NLLNYKTSNIETLS	148
QY	98	PNPKSDFKLLGSDV-FANLVAAL-ENYSFENV	-----LVVDLGT	143
DB	149	NNFYLMDDYLYNKMYKLDLNNPNNYNNFNINYGKLNKIVTLDLNNNEFYDYLSGLI	208	
QY	144	RDGIL--GGI	-----INSGLIN--FNSLLDNAYLIKKEPISTPNN	181
DB	209	EGDGYIGPGGITITNHANDVLTITFKIKNSILVEKWMDTLKDNPYFVNAFSINKTN	268	
QY	182	LLERTSGSVNSGLFYQYK	-----YL-----IEGVYRDIK	211
DB	269	LAKEKIFNIYKNLYSDYKINNNHPIYNNKLNKLPKIKNI-MDKNNYVLAGFTAA	327	
QY	212	-----QMKKKFNLIITGGNADLLIISLIEI	-----LVEGVRIILGNS--ID	257
DB	328	DGSLSSMYNPKDITLLFKNNRPSYVISOVETRKRELIYIOESPDLSISNVKVKYGNRKLKD	387	
QY	258	FK 259		

Db 388 FK 389

## RESULT 2

US-09-140-466-2

Sequence 2, Application US/09140466

Patent No. 6268160

GENERAL INFORMATION:

APPLICANT: CLOUGH, BARBARA

APPLICANT: PREISER, PETER

APPLICANT: WILSON, ROBERT

TITLE OF INVENTION: AN EF-TU PROTEIN ENCODED ON THE PLASTID DNA OF THE

TITLE OF INVENTION: MALARIA PARASITE AND PROTEIN SYNTHESIS INHIBITORS

TITLE OF INVENTION: EFFECTIVE AS ANTI-MALARIAL COMPOUNDS

FILE REFERENCE: N68837B GCW PJC DP

CURRENT APPLICATION NUMBER: US/09/140,466

EARLIER FILING DATE: 1998-08-26

EARLIER APPLICATION NUMBER: US 60/056,246

EARLIER FILING DATE: 1997-08-28

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 410

TYPE: PRT

ORGANISM: Plasmodium falciparum

US-09-140-466-2

Query Match

Best Local Similarity 7.5%; Score 99.5; DB 4; Length 410;

Matches 55; Conservative 39; Mismatches 68; Indels 77; Gaps 12;

QY 9 LIIDIGNTSIAPALFKD---NOVNL--FIKMTN-LMLRYDEVISFFEFDFNVNKFVI 62

Db 122 LIKQIGIKNIIFLNKEDLDDVELIDFKLEVNELLKY-----NFDLNYTHILT 172

QY 63 SSVVPLNETFKNVIFSPFKIRPLFGFDLNDLTFNPNKSKFLGSDVFA-----NIVA 118

Db 173 GSALNVINIQRN-----KQYELKSNITQKLNLIQ 205

QY 119 AIENYSFENVLVVD--LGTACTIFAVSRQDGLGGIINSGLTINFNSLDNLAYLIKFFI 176

Db 206 IIDNIIPRKINDYFLMSIEDVFSITGRGTVTGKIEG-CINLNDIEE---ILKFEK 260

QY 177 STPNLLERTTSGSVNSGLFYQYKLYIEGVYDIQMYKKKNLIITGNADLILSLIE 235

Db 261 SSPN-----LTT-----VIGL-----EMFKQLTQAGSDGVNGLLRNIQ 295

## RESULT 3

3-08-477-451-8

Sequence 8, Application US/08477451

Patent No. 5928865

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello

TITLE OF INVENTION: Helicobacter Pylori CagI Region

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: CA

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,451

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0335.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-601-2708

TELEFAX: 510-655-3542

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 3200 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-477-451-8

Query Match

Best Local Similarity 7.3%; Score 96.5; DB 2; Length 3200;

Matches 57; Conservative 52; Mismatches 109; Indels 73; Gaps 11;

QY 10 IIDIGNTS-IAFALFKDQVNL-FIKMKTNLMRLRYDEVISFFEFDFNVNKFVSSVVP 67

Db 1059 ILSVSDCSCVSIISLFDQKNLIFLAVSLSCFHFIFVAAAFALQFFLY---PFLMLSEF 1114

QY 68 ILNETFKNVIFSPFKIRPLFGFDLNDLTFNPNKSKFLGSDVFNALVAAIENYSFEN 127

Db 1115 ---REFKIFLSPFHIITACVFSVSFTIEVSFSNOFFVAAQKIFDKLISFLMGQSFSG 1170

QY 128 VLVDLGTACTIFAVSRQ-----DGLGG-----IINSGPL 158

Db 1171 FILIPPYSDSEIIIRQKDSYLVRAFMGNAINRAYTLIGLGGVGVFEIRIINKHHA 1230

QY 159 INFNSLLDNLAYL-----IKKFPISITPNN-----LLERTTSGSVNSGLFYQYK 200

Db 1231 INFATLKSNIPLPISFCFPIIFESVAQIPITKLSNMFVFRNKSNIIPGV-----1285

QY 201 YLIEGVYDIQMYKKKNLIITGGNADLILSLI-----EIFIFNIHLT 245

Db 1286 -ISLRVLEIFNLSEFFLLFL---AQVVFSLICKKFKHEIYFVLDFLT 1331

## RESULT 4

US-08-686-968C-4

Sequence 4, Application US/08686968C

Patent No. 6221361

GENERAL INFORMATION:

APPLICANT: Cochran, Mark D.

APPLICANT: Junker, David E.

TITLE OF INVENTION: Recombinant Swinepox Virus

FILE REFERENCE: 39119-H/JML

CURRENT APPLICATION NUMBER: US/08/686,968C

CURRENT FILING DATE: 1996-07-25

NUMBER OF SEQ ID NOS: 231

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 4

LENGTH: 493

TYPE: PRT

ORGANISM: Swinepox virus

US-08-686-968C-4

Query Match

Best Local Similarity 7.1%; Score 95; DB 4; Length 493;

Matches 57; Conservative 51; Mismatches 100; Indels 68; Gaps 14;

QY 6 LSELIIDIG-NTSIAFALFKDQVNLFIKMTNLMRLRYDEVISFFEFDFNVNKFVSSVVP 60

Db 159 LVLELLQKGVNDVNIHNNWFYFNLHCYIKKNYN-RINMDLIKIMONGFTINENKFKST 217

QY 61 FISSVVPIL-NETFKNVIFSF-----FKIKPLFGFDLNDLTFNPNKSKDKFL 107

Db 218 FLDLVLISLIDSKNDSNVVDFFILKYIDINEKNIFDTFLYCSVDANNE-----KMSYL 271

QY 108 LGSDFANLV-----AAIENYSFENVLV-----VDLGTACTIFAVSRQDGLGGI 152

Db 272 LKNADPNITVFGETCILTAINNN-KNLYKLLNYDIDINTQNTLFKLEQD-----I 325  
QY 153 INSG-PLINFLNLDN-----AYLIKPPISTPNNLLERTTSGSVNSGLFYQYKYL 203  
Db 326 INSTIDYNNVNNVKKHEFIKFLAYIVKRYEKNGILFLDYPTLG-----EYFV 375  
QY 204 EGVYRDIKOMYKKFNLIITGNGADLILSLIEIFI 239  
Db 376 KFIDTCMEIFEMKSD---KAGNTDIYSIITNKYI 408

## RESULT 5

US-09-134-001C-2927  
; Sequence 2927, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2927  
; LENGTH: 284  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2927

Query Match 6.9%; Score 92.5; DB 4; Length 284;

Best Local Similarity 22.3%; Pred. No. 0.055;

Matches 60; Conservative 35; Mismatches 83; Indels 91; Gaps 13;

QY 8 ELIIDIGNTSTAFALFXDNQVNLTKM-----KTNLML-----RYDEVYSFFFEEN 52

Db 7 KLIGEVKMHYIKFTESKDN-TKLYMKVNDIQDANKANIIAHGVAEHLRDYEITAY----- 61

QY 53 FDFNVNKFIVSSVPIINLETKNVIFSFKKIKPLFIGFD-LNYDLTFNPKYSDK--FLLG 109

Db 62 -----LNEA-----GFSVIRYDQRGHGRSEGRAFYSN 89

QY 110 SDVEANIVAAIENY--SFENVLVV---DLGTACTIFAVSRQDGLGIIINSGLPINFS 163

Db 90 SNEIVEDLDAIINVKSNFEGKVVYLIGHSMGTYTWTLYGTHKHPNTVNGIITSGALTRYNN 149

QY 164 LLDNAYLIKPPISTPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKFNLIIT 223

Db 150 KL-----FQNPDRNISPD-----YIENNLSEGVCSDEVMEKYKLD----- 186

QY 224 GGNADLILSLIEIFINHLTVGVRIL 252

Db 187 ----DLAKQISMGLVFSI---MDGVRYL 208

## RESULT 6

US-09-134-001C-3626  
; Sequence 3626, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3626

; LENGTH: 331

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3626

Query Match

6.9%; Score 92; DB 4; Length 331;

Best Local Similarity 21.4%; Pred. No. 0.078;

Matches 49; Conservative 31; Mismatches 85; Indels 64; Gaps 8;

QY 32 IKMKTNLMLRYDEVYSFPE-----ENEDFENVNK--VFISVVVPIINLETFKNV 76

Db 6 VRMSNNITVYGEVPELVKKEIVNDYLGQEIIDFENVKYNLYESDLTPIIEITLTP 65

QY 77 IFS-----FFKIKPLFIGFDLNYDLTFNPKYSDKFLGSDVFANLVAAIENYSFENVLVV 131

Db 66 PFSNKKAIWKNYSYVFTGEKFSKDLNHN---SDE-----LIKFLKEYDGENLIIF 112

QY 132 DLGTACTIFAVSRQDGLGIIINSGLPINFNSLIDNAYLIKPPISTPNNLLERTTSGSV 191

Db 113 EV-----YQPKLDER---KKLTKLKRNAOLKKIBOMS 142

QY 192 NSGLFYQYKYLIEGVYRDIKOMYKKFNLIITGNGADLILSLIEIFIF 240

Db 143 EKELKHVKNLNNYKDIKQDALELF-IELGTGVNIVNYSQLEKILIF 190

## RESULT 7

US-09-134-001C-5409

; Sequence 5409, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5409

; LENGTH: 574

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5409

Query Match

6.9%; Score 91.5; DB 4; Length 574;

Best Local Similarity 30.5%; Pred. No. 0.2;

Matches 36; Conservative 20; Mismatches 41; Indels 21; Gaps 7;

QY 19 AFALFKDQVNLFIKMTNMLRYD---EVYSFEENFDFENVKVFISVVVPIINLETFK 74

Db 435 ATVLFK---MHPFVRKNLNPAYISKYFLDIGNRVNDVLFITDILISYSSLIYE--- 488

QY 75 NVIFSFKKIKPLFIGFDL-NYDLT---FNPYKSDKFLGSDV--FANLVAAIENYSPE 126

Db 489 ---FSVFKKPMFLYAFDLEDIYTRDFYEPYET--FVPGKIVKTFDELILALENDPFE 541

## RESULT 8

PCT-US94-00198-4

; Sequence 4, Application PC/TUS9400198

; GENERAL INFORMATION:

; APPLICANT: Schering Corp.

; TITLE OF INVENTION: RAS Associated GAP Proteins

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schering Corp.

; STREET: 1 Girard Farms



; GENERAL INFORMATION:  
 ; APPLICANT: Kupke, Thomas  
 ; APPLICANT: Gotz, Friedrich  
 ; APPLICANT: Kempfer, Christoph  
 ; APPLICANT: Jung, Gunther  
 ; TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
 ; TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD  
 ; NUMBER OF SEQUENCES: 70  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 ; STREET: 1100 New York Avenue, Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/645,193B  
 ; FILING DATE: 13-MAY-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Esmond, Robert W.  
 ; REGISTRATION NUMBER: 32,893  
 ; REFERENCE/DOCKET NUMBER: 0652.1540000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600  
 ; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 19:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 275 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; US-08-645-193B-19

Query Match 6.5%; Score 86; DB 2; Length 275;  
 Best Local Similarity 20.5%; Pred. No. 0.26;  
 Matches 55; Conservative 49; Mismatches 94; Indels 70; Gaps 13;  
 QY 4 PLLSELIIDGN-----TSIAPALFKDQVNLFIKMTNMLRYDEV--YSPFE 50  
 DB 28 PILIIVILPIGNIKRVSSKSQEATKLSYSNRLSTIKLIKLTSTYNIKIRNYTLK 87  
 QY 51 ENFDENVKV-FISSVVPILNETKFNIFSEFKIKPLFGLFDLNYDLTFNPKSKDKFLLG 109  
 DB 88 NIFDIELHKIKVLSFFPEIMNLILFINIFGI-----LFLGYLH-----ENNMMKS 133  
 QY 110 SDVEA-----NLVAAIENYSFENVLVVDLGTACTIFAVSRDGDILGGLINSGLPIN 160  
 DB 134 GDMFAVLYLQIINPIVSITSYTE-----VQRAIG-----SSDRILK 172  
 QY 161 FNSLLDNYALKPKPIISPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDKQYKKFNL 220  
 DB 173 INKEPEVLTIK-----TTYNNFVKMEINDLN--FTKDNKQILNISLHLKGY--IYNI 224  
 QY 221 IITG--GNADL--ILSLIEIEFIFNIHL 244  
 DB 225 IGESGCGKSTLLNLAGLNTGYTNIYL 252

RESULT 11  
 US-09-134-001C-3433  
 ; Sequence 3433, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3433  
 ; LENGTH: 299  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-3433

Query Match 6.5%; Score 86; DB 4; Length 299;

Best Local Similarity 23.8%; Pred. No. 0.3;  
 Matches 67; Conservative 50; Mismatches 87; Indels 78; Gaps 20;  
 QY 32 IKMKTNMLRYDEVYSFPEENFDVNVK-----VFISVVVPILNETKFNIFSF 80  
 DB 15 VKLKNLISELKVMETIIMH-----NINKWLLVPLLMMVIFI--IIPVI-----LLYFVSF 64  
 QY 81 FKIKPLFGLFDLNYDLTFNPKSKDKFLGSDVFAANLVAIE-----NYS-FENV 128  
 DB 65 IDIHGF-SF-TNYEQVFST-RYLKMFIDSIIWAALITWITLIIISYPAAFFISYSRFQNI 121  
 QY 129 LVVDLGTACTIFAVSRQ---DGILG--GIINSG-----PLIN--FNS---LLDNAYLIK 172  
 DB 122 LLMLLIIPWINLLKTYAFIGLLGHGDGVINQALHFQIPKLNLLFTSGAFLLVASVYII 181  
 QY 173 KPPI-----STPNLLERTTSGSVNSGLFYQYKYL-----EGVYRDKQYKKFNF 219  
 DB 132 PFMILPIFNSMKAIPNNILQ--ASNDLGASTTTTRKVIPLTREGIKTGVTTFIPALS 239  
 QY 220 L-----IITGGNADLILSLIEIEFIFNIHLTVEGVRLGNSI 256  
 DB 240 LFMITRLIAGNKVINVGTAIEQF-----LTQNYG-LGSTI 275

## RESULT 12

US-09-134-001C-5654  
 ; Sequence 5654, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5654  
 ; LENGTH: 425  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 ; US-09-134-001C-5654

Query Match 6.3%; Score 84.5; DB 4; Length 425;  
 Best Local Similarity 21.9%; Pred. No. 0.72;  
 Matches 37; Conservative 39; Mismatches 58; Indels 35; Gaps 9;

QY 18 TAFALFKDQV----NLFKMTNMLRYDEVYSFPEENFDVNVKVFISVVVPILNETF 73  
 DB 70 VAFSIDKIEQIYELNQLFTYQYENLKIMNKCSVSDYNVOQIFNLNKS-----REIY 122  
 QY 74 KNVIESFFKIKPLFGLFDLNYDLT--FNPKSKDKFLGSDVFAANLVAATENYSFENVLVV 131  
 DB 123 RNIL----KIQNIELEYKIGLDISCMFN-----DTAQFASLASQIKRLAFD-YLYI 168





```

;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-2

Query Match 6.3%; Score 84.5; DB 1; Length 1536;
Best Local Similarity 22.7%; Pred. No. 4.7; 77; Indels 99; Gaps 16;
Matches 64; Conservative 42; Mismatches 77; Indels 99; Gaps 16;

QY 23 FQDNQVNL-----FIRKNTLMRLRYDEYISFFENFDNKNKVFISVVPIILNET-- 72
   | : | : | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 601 FRENNSVNLGTGSLQFTTARTN--KY-AITNKFEGTLNIS-CKVNISMVLP-KNESGY 654
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 73 --FKNVIFSPFKIKPLFI-----GFDLNYDLTFNPK 102
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 655 DKFKG--RTYWNLTSLNVSEGEFNLTIDSRGSDSAGTLQPYNLNGISFNKOTFNVER 712
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 103 SDR--FLIGSDVFNVAATENYSENVLVVDLTAC--TIFAVSRQDGIILGILNSGPL 158
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 713 NARNVFDIKAPIGINKYSSLNYSFNGNISVSGGSDVDFTLASSNNVQTFGVVINS--- 769
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 159 INFNSLLDNAYLIKFPPISTPNILLERTTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKF 218
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 770 -----KYFNVSTGSSLRFK-TSGSIKYG-----FSIE-----KDL 798
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 219 NLIITGNADLIILSLIE-----FIFNHLTVGVRI 251
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 799 TLNATGNN-----ITLQVEGTGDMIGKGIIVAKKNITPEGGNI 836
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

search completed: June 24, 2003, 21:49:25  
Job time : 9.65383 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.55691 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453a-11

Perfect score: 1331

Sequence: 1 MNKPLISELIIDIGNTSIAP.....HLTVEGVRIILGNSIDFKFVN 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231.5	17.4	233	1 YACB_BACSU	P37564 bacillus su
2	120.5	9.1	476	1 RF3_SACBA	P05512 saccharomyc
3	109.5	8.2	2136	1 YCF2_MARPO	P09975 marchantia
4	108	8.1	311	1 PRMA_CIOAB	P45558 clostridium
5	106.5	8.0	519	1 YC82_ASTIO	P58145 astasia lon
6	102	7.7	1024	1 RPOB_PLAFA	P21421 plasmodium
7	100	7.5	355	1 Y198_RICPR	Q9zd47 rickettsia
8	99.5	7.5	516	1 Y198_RICPR	P30397 euglena gra
9	99	7.4	971	1 YQ1A_CAEEL	Q9281 caenorhabdi
10	96	7.2	534	1 Y478_RICPR	Q9zd66 rickettsia
11	95.5	7.2	773	1 SYFB_CAMJE	Q9pp35 campylobact
12	95	7.1	971	1 SEC5_YEAST	P89102 saccharomyc
13	93.5	7.0	507	1 MVIN_RICPR	Q9zcw4 rickettsia
14	93.5	7.0	1783	1 Y468_MYCGE	Q49460 mycoplasma
15	93	7.0	482	1 ANFR_AZOVI	P16267 azotobacter
16	93	7.0	1149	1 A8A1_BOVIN	Q29449 bos taurus
17	92.5	6.9	269	1 FLIR_BORBU	Q44907 borrelia bu
18	92.5	6.9	2054	1 YCF2_PINTH	P41653 pinus thunb
19	92	6.9	518	1 PAC2_YEAST	P39937 saccharomyc
20	92	6.9	1149	1 A8A1_MOUSE	P70704 mus musculu
21	92	6.9	1164	1 A8A1_HUMAN	Q9y2q0 homo sapien
22	91.5	6.9	648	1 GCP4_DROME	Q9vk47 drosophila
23	91.5	6.9	1442	1 DPO3_UREPA	Q9pqb4 ureaplasma
24	91	6.8	322	1 GPT_SULSO	P96000 sulfolobus
25	91	6.8	363	1 V363_ASPE7	P23164 african swi
26	90.5	6.8	504	1 Y795_METJA	Q58205 methanococc
27	90.5	6.8	575	1 RPOC_PLAFA	P21422 plasmodium
28	90.5	6.8	602	1 EX5A_BUCAI	P57530 buchera ap
29	90.5	6.8	1296	1 BXG_CIOBO	Q60393 clostridium
30	90	6.8	708	1 PTA_BUCAI	P57273 buchera ap
31	90	6.8	795	1 SYFB_BOVIN	P57230 buchera ap
32	89.5	6.7	279	1 PPNK_BORBU	Q51291 borrelia bu
33	89.5	6.7	380	1 IPYR_PLAF7	O77392 plasmodium

34	89.5	6.7	546	1 NTP1_HAEPV	O37319 heliothis a
35	89	6.7	339	1 RMAR_CANGA	P21358 candida gla
36	89	6.7	361	1 V227_FOWPV	Q9j508 fowlpox vir
37	89	6.7	623	1 PPID_BUCAI	P57550 buchera ap
38	89	6.7	898	1 YMV6_YEAST	Q04748 saccharomyc
39	89	6.7	3079	1 IRA2_YEAST	P19158 saccharomyc
40	88.5	6.6	229	1 Y066_BORBU	O51093 borrelia bu
41	88.5	6.6	259	1 IF2C_GALUS	Q08810 galdieria s
42	88.5	6.6	461	1 TRPC_BUCSC	Q44603 buchera ap
43	88	6.6	813	1 VGLR_HSVMD	P36336 marek's dis
44	88	6.6	1250	1 BXE_CIOBO	Q00496 clostridium
45	87.5	6.6	354	1 PON3_HUMAN	Q15166 homo sapien

## ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Bourris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blauchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,  
Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivalet C., Rocha E., Roche B., Rose M., Sadale Y.,  
Sato T., Scanlon E., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D26185; BAA05305.1; -  
 CC EMBL; Z99104; CAB11846.1; -  
 CC Subtilist; BG10133; YACB.  
 CC InterPro; IPR004619; Baf.  
 CC Pfam; PF03309; Bvg\_acc.factor; 1.  
 CC TIGRfams; TIGR00671; baf; 1.  
 CC KW Hypothetical protein; Complete proteome.  
 CC SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;  
 CC -----  
 CC Query Match 17.4%; Score 231.5; DB 1; Length 233;  
 CC Best Local Similarity 26.8%; Pred. No. 3.4e-11;  
 CC Matches 59; Conservative 48; Mismatches 92; Indels 21; Gaps 5;  
 CC -----  
 CC YJ 9 LIIDIGNTSTAPALFKDQNVNLIKKTNMLRYDEVYGFEEFNDFN-----VNKVF 62  
 CC Db 3 LVIDIGNTVLGVHDGLEYHWRITSRHKTEDFGMLRSLFDFHSLMFEQIDGII 62  
 CC QY 63 SSVVPIILNETFKNVIFSFKKIKPLFI-----GFDLNYDLTFNPYKSDKFLIGSDVFANL 116  
 CC Db 63 SSVVPIIMFALERMCYKYPHIEPQIVGPMKGTGLNKYD---NPK-----VGADRIVA 114  
 CC QY 117 VAAIENYSFENVLVDLGTRACTIFAVSRDGLGGIINGSPILNFINSLDNAYLIKFFPI 176  
 CC Db 115 VAAIHLVG-NPLIVDFGTATTVCYIDENKQYMGGAIPGITTISTEALYSRAAKLPRIE 173  
 CC QY 177 STPNLLERTSSVNSGLFYQYKYLIEGYRDIKQYK 216  
 CC Db 174 TRPDNIIGKNTVSAMSGILFYGVQGVGIVKRMKQAKQ 213  
 CC -----  
 CC RESULT 2  
 CC RF3\_SACBA  
 CC ID RF3\_SACBA STANDARD; PRT; 476 AA.  
 CC AC P05512;  
 CC DT 01-NOV-1988 (Rel. 09, Created)  
 CC DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Maturase-like RF3 protein.  
 CC GN RF3.  
 CC OS Saccharomyces bayanus (Yeast) (Saccharomyces uvarum).  
 CC PE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC OX NCBI\_TaxID=4931;  
 CC RN [1]  
 CC RE SEQUENCE FROM N.A.  
 CC RP MEDLINE=87280035; PubMed=2440860;  
 CC RA Seraphin B., Simon M., Faye G.;  
 CC RT "The mitochondrial reading frame RF3 is a functional gene in  
 CC Saccharomyces uvarum.";  
 CC RL J. Biol. Chem. 262:10146-10153(1987).  
 CC SC -1- SIMILARITY: THE RF3 PROTEINS ARE RELATED TO THE MATURASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J03300; AAA32166.2; -  
 CC FIR; A28439; A28439.  
 CC InterPro; IPR001982; Endonuc\_LAG/HNH.  
 CC -----

```
DR InterPro: IPR003959; AAA_ATPase_centre.
DR Pfam: PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match      8.2%; Score 109.5; DB 1; Length 2136;
Best Local Similarity 23.4%; Pred. No. 1.3;
Matches 80; Conservativity 55; Mismatches 98; Indels 109; Gaps 21;

QY 6 LSELIIDIGNT-----STAFALF-----KDNQVNLFIKMKTNLMRLRYDEV 45
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 INOLKIDIKSNKYNKSGFVFLAFCEKLLFEVFLSKPNNNL--QMKLNCL----- 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 46 YSFEENFDN-----VNKVFISVVPIILNETFNKVFSPFKIRPLFGFDLN 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 -----ENFSFDIFCILNKKLPWNKKTFKN--LQNFNEDSKLIESFFLLKIKGNLYFKN 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 94 Y--DLTFNPKSD-----KF--LLGSDVFANLVAAIEN--YSFENVLVVDLGTACTIFAV 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 YIEFTWQSYKDKCLDFKFNELNNESEIYKIEELFSDIYKFSYIIEGKKSTIIKQ 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 143 SRQGIILGGIINSOPLINSL-----LDNAYLIKKFPI--STPNNL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 SFNNIYYKKLNS--IFNENTIFYDSNNLLFDWLKKNYINNKPFKSLIYSISNOF 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 183 L---ERTTSGVNSGLFYQYKYLEGVYRD--IKMKYKKFNLIITGGNADL----- 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 ILFFKQKNSKFN-----KNLVKNSKDVITNVSFENKTEINNFSSKIYAFPEILS 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 233 LIETB--FIENHLTVEGV-----RILGNSI-----DFKEVN 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 INEIDNKEVIN--KISLKNINKKKQKRYLNLIKSSDNPRFIN 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
PRMA_CLOAB
ID PRMA_CLOAB STANDARD; PRT; 311 AA.
AC P45558;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable methyltransferase (EC 2.1.1.-).
GN PRMA OR CAC1284
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
N [1]
P SEQUENCE FROM N.A.
AC STRAIN=ATCC 824 / DSM 792 / VRM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
N [2]
RP SEQUENCE OF 1-298 FROM N.A.
RX MEDLINE=94123950; PubMed=7507453;
RA Behrens S., Narberhaus F., Bahl H.;
RT "Cloning, nucleotide sequence and structural analysis of the
RL FEMS Microbiol. Lett. 114:53-60(1993).
N [3]
RP REVISIONS.
RA Behrens S.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: STRONG, TO ENTEROBACTERIAL RIBOSOMAL PROTEIN L11
CC -1- SIMILARITY: STRONG, TO OTHER METHYLTRANSFERASES.
CC or send an email to license@isb-sib.ch).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE007641; CAA79255.1; -
CC EMBL: X69050; CAA48793.1; -
CC InterPro: IPR004498; Pfam.
CC InterPro: IPR000051; SAM_bind.
CC TIGRfams: TIGR000406; Pfam; 1.
CC Heat shock; Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 311 AA; 35028 MW; 2D71786B3C974922 CRC64;

Query Match      8.1%; Score 108; DB 1; Length 311;
Best Local Similarity 20.9%; Pred. No. 0.17;
Matches 58; Conservativity 52; Mismatches 87; Indels 80; Gaps 15;

QY 8 ELIIDIGNTSIAFALFKDNQVNLFIKMKTNLMRLRYDEVYFFFEENFD-----FNVN---K 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 ESSLNVKOGAVIKAYKDDH-----NFDESVKYIEESIDKLSFEGINKGEGK 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 VFISVVPIILNET--FKNVIFSPFK-----IKPLFTGF--DLNYDLTFNPKSKDPL 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 VEVNRY-----NETDWNENKKYKPKIGARIVVVKPLWEETPKDYELNMDPGMAFGT 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 109 GSDVFANL--VAIENYSFENVLVVDLGTACTIFAVS-----RDQGIILGGIINSGLP----- 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 GTHETTRMCIOALERTYVNEDAEVFDIGTSGIILAAAKLNNAKKVLGVLDLSDVAVKAAKE 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 159 -INFNSLDNAYLIKKFPISTPNLLERTTSG-----GNLME--VVGKADIIVANIADVINILIPDINKFLKTD 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 NIQYNN--VNNIELH-----IKMKYKKFNLIITGGNADLI 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 199 KYLIEGVYRD-----IKMKYKKFNLIITGGNADLI 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 GYFISSGIILKRAEDVIENLKNKFEIIEVNNQGEWI 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
YC82_ASTLO
ID YC82_ASTLO STANDARD; PRT; 519 AA.
AC P58145;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 65.0 kDa protein ycf82 (ORF519).
GN YCF82.
OS Astasia longa (Euglenophyceae alga).
OG Chloroplast.
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.
OX NCBI_TaxID=3037;
N [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CCAP 1204-17a;
RX MEDLINE=21080522; PubMed=11212895;
RA Cockel G., Hachtel W.;
RT "Complete gene map of the plastid genome of the nonphotosynthetic
euglenoid flagellate Astasia longa."
RL Protist 151:347-351(2000).
CC -1- SIMILARITY: BELONGS TO THE YCF82 FAMILY. SOME SIMILARITY TO
CC GROUP II INTRON MATURASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



Query Match 7.5%; Score 100; DB 1; Length 355;  
 Best Local Similarity 25.7%; Pred. No. 0.85;  
 Matches 66; Conservative 38; Mismatches 87; Indels 66; Gaps 16;

QY 2 NKPL-----LSLIIDIGNTSTAFALFKDQ-----VN--LFTKMTNLMRYDEVYS 47  
 DB 127 NKPLYKQSDSLKELISIKLNKTKLSIFKDNQEIFKINDLAFLIKHNLA-----178

QY 48 FFEENFDNPNKVFSSVVPILNETPKNVIFSFKKPLFIFGDLNLYDLTNPYKSDRFL 107  
 DB 179 -SQENISLNNHYSE-KDILN--FKNA-----NLDMATSPFKANGK--218

QY 108 LGSDFANLVAIENYSF--ENVLVVDLGTACTIFAVSRQDGLGGIINSGLINFNLSLL 165  
 DB 219 -DAAILENL--NIERFIFCTDNESKVNLTNGTQFFANKLPKGL-----SFELENYNSIV 270

QY 166 D-----NAVLIKKFFIST--PNLLERTTSGSVNSGLFYQYKYLIEGVYDIKQMYKKKNL 220  
 b 271 DKLPLNSILFKKTKTKTKIAKAMKNTSDEQLNTD-----KNDTNSVYNNIK---NAKFDI 322

QY 221 IITGGNADL--ILSLIEI 236  
 DB 323 AFSDKGINSGMNLLEL 339

## RESULT 8

YC82\_EUGGR  
 ID YC82\_EUGGR STANDARD; PRT; 516 AA.  
 AC P30397;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 64.3 kDa protein ycf82 (ORF516).  
 GN YCF82.  
 OS Euglena gracilis.  
 OC Chloroplast.  
 OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.  
 OX NCBI\_TaxID=3039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Z;  
 EX MEDLINE=93347989; PubMed=8346031;  
 RA Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,  
 RA Orsat B., Spielmann A., Stutz E.;  
 RT "Complete sequence of Euglena gracilis chloroplast DNA.";  
 RL Nucleic Acids Res. 21:3537-3544(1993).  
 C -1- SIMILARITY: BELONGS TO THE YCF82 FAMILY. SOME SIMILARITY TO  
 C GROUP II INTRON MATURASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X70810; CAA50104.1; -  
 DR EMBL; Z11874; CAA77921.1; -  
 DR PIR; S26088; S26088.  
 DR PIR; S34525; S34525.  
 KW Chloroplast; Hypothetical protein.  
 SQ SEQUENCE 516 AA; 64333 MW; E2079EF48423490A CRC64;

Query Match 7.5%; Score 99.5; DB 1; Length 516;  
 Best Local Similarity 20.0%; Pred. No. 1.4;  
 Matches 42; Conservative 32; Mismatches 61; Indels 75; Gaps 8;

QY 27 QVNFILKMTNLMRYDEVYSFEEENFDNPNKVFSSVVPILNETPKNVIFSEFKIKPL 86  
 DB 227 QENFFISYKVVYLFYIDKFFPKFPYDF-----FKIISFESLGI 271

QY 87 FIGFDLN-----YDLTFNPKSKDKFLGSDVFANL--VAALIENYSFENVLVVDLGTACTIF 140  
 DB 272 ILINININSYRPFETVFNFFIDFLRFGSNILHINKKHIOFYKLSLKMIVKMLPKKSVF 331  
 QY 141 AVSRQDGLGGIINSGLINFNLSLDNAVLIKKFFISTPNLLERTTSGSVNSGLF-----196  
 DB 332 -----FLV-----NLLNKKILDCNLNRGFFCFNN 354

QY 197 -----YQYKYLIEGVYDIKQMYKKK 217  
 DB 355 NLFLELDLYLRLL-----WRYIKLHSRK 380

## RESULT 9

YQIA\_CAEEL  
 ID YQIA\_CAEEL STANDARD; PRT; 971 AA.  
 AC Q09281;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Hypothetical 111.8 kDa protein C45G9.10 in chromosome III.  
 GN C45G9.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Bentley D., Waterston R.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; U21323; AAA62553.1; -  
 DR WormPep; C45G9.10; CE01851.  
 KW Hypothetical protein.  
 SQ SEQUENCE 971 AA; 111772 MW; 6AA5EB10E2BB7693 CRC64;

Query Match 7.4%; Score 99; DB 1; Length 971;  
 Best Local Similarity 23.1%; Pred. No. 3.3;  
 Matches 56; Conservative 37; Mismatches 83; Indels 66; Gaps 12;

QY 2 NKPLLSLELIIDI-----GNTSIAFALFKDNVNLFIRKMT-NLMURYDEVYSFFENFDN 56  
 DB 25 NNGLLTDKILKITSHPGNVSV-----WDINEVIVFGEQSGKLSLILPKQFSIKLEKVDRT 80  
 QY 57 VNKVFSSVVPILNETPKNVIFSFKKIKPLF--IGFDLN-----Y 94  
 DB 81 FLNVVF--LKPIDDDYFHNVIETVSKEMWPSLQFDLQSIHLKIVDRSQSQKDLILLY 138  
 QY 95 DLTFFNPKSKDKFLG--SDVFANLVAALIENYSFENVLV-----VDLGTACTI---139  
 DB 139 DLRAEQQTSKILLAPNKRDPKQRPVMDTIHSQVVFQVQNAMNRSHVSSGSRCTAFL 198  
 QY 140 -----FAVSRQDGLGGIINSGLINFNLSLDNAVLIKKFPSTPN-----NLERT 186  
 DB 159 LNTNGDFENVYTSKNDEI--RILNYTKIINAN---EYAILVEFETLKPCKAEFNVVQKQ 253

187 TS 188  
 254 IS 255

## RESULT 10

Y478\_RICPR  
 ID Y478\_RICPR STANDARD; PRT; 554 AA.

```

AC O92D66;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP478 precursor.
GN RP478.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RT "The genome sequence of Rickettsia prowazekii and the origin of
  mitochondria."
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ235271; CAAL4933.1;
DR Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 554 HYPOTHETICAL PROTEIN RP478.
SQ SEQUENCE 554 AA; 64341 MW; EC2C5E9AE2F373E CRC64;

Query Match 7.2%; Score 96; DB 1; Length 554;
Best Local Similarity 23.3%; Pred. No. 2.9;
Matches 60; Conservative 28; Mismatches 81; Indels 82; Gaps 14;

QY 3 KPILSELIIDIGNTSIAFALDKDNQVNLPIKMTNLMRYDEVISFFERNFVNKVF 62
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 264 KDIADKLTIKVNAKNSNNF-ENNMLLYKGLN-----DFGNSNIHL 306
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 SSVVPIINTEFNKVFSPKIRPLF-IGF-----DLNYDLTFNPKSKDKFLGSD--VFAN 115
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 307 S-----IESQFKRPFGTIGFLEFKKNYD-----KOTYLLKFSDNKIYN 347
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 LVAAIENYSFENVVVDLGTACTIFAVSRQDGLIGGIINSGLPLNFNSLLDNYALLKFP 175
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
  348 -----FNELVY-----ILNNKPYNFSEILEDPRYPF-NFN 377
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 ISTPNLLERTTSGSVNSGLFYQ-----YKYLIEGVYRDIKQYKKFNLIITGGNADLIL 231
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 378 INLVPE-LKKLTRVQINTLSYNTSGFNITNETIINDLKDSYTKG---LIIVNYSKII 433
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 SLIEIEFIFNI 242
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 434 EILSF-YIYG 443

RESULT 11
SYTB_CAMJE STANDARD; PRT; 773 AA.
AC Q9PP35;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
  DE tRNA ligase beta chain) (PheRS).
GN PHET OR CJ0896.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.

OX NCBI_TaxID=197;
RN [1]
RT SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
  Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
  Jegerl M.A., Kariyshev A.V., Moule S., Pallen M.J., Penn C.W.,
  Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
  Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
  reveals hypervariable sequences."
RL Nature 403:665-668(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) -> AMP +
  CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by
  CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
  CC CHAIN FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL139076; CAB73154.1;
DR InterPro; IPR005146; B3_4.
DR InterPro; IPR005147; B5.
DR InterPro; IPR005121; Fdx-AnticB.
DR InterPro; IPR003583; HH1_1.
DR InterPro; IPR004532; Phet_bact.
DR InterPro; IPR002547; tRNA_bind.
DR Pfam; PF01588; tRNA_ACB; 1.
DR Pfam; PF03147; FDX_ACB; 1.
DR Pfam; PF03483; B3_4; 1.
DR Pfam; PF03484; B5; 1.
DR SMART; SM00278; HH1; 1.
DR TIGRFS; TIGR00472; phet_bact; 2.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
  KW Complete proteome.
SQ SEQUENCE 773 AA; 87250 MW; 21FE593C10AEFEF CRC64;

Query Match 7.2%; Score 95.5; DB 1; Length 773;
Best Local Similarity 22.3%; Pred. No. 4.7;
Matches 48; Conservative 32; Mismatches 70; Indels 65; Gaps 9;

QY 69 LNETFKNVIFSPFKIRPLFIFGLDNLDTNPNYKSKDKFLGSDVFANLVAIEN----- 122
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 478 LNSAYKNYI-EFLNLR-----KRAVASGVFESLHYVLDNGEELKR 516
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 YSFENV-----LVVDLGTACTIFAVSRQDGLIGGIINSGLPLNFNSLLDNYALLKFP 175
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 517 LGFDSYKVLKLIINPITAEALNLT-----LNLHLNAAASNAKNS-----KKIKLFE 564
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 ISTPNLL--LERTTSGSVNSGL-----FYQYKYLIEGVYRDIKQYKK 217
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 565 LGAVFNVNQELNRIAFIHSGLKEAKISKAKPESVQFYDFELDIKRIIGDFK-LKSSK 623
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 FNLITGGNADLILSLIEIEFIFNIHLTVEGVRL 252
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 624 YNLSPEQADIYLSIDIKVGFIGRLHLKIENERDL 658

RESULT 12
SEC5_YEAST
ID SEC5_YEAST STANDARD; PRT; 971 AA.
AC P89102; Q04128; Q03775;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```



DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Exocyst complex component SEC5.  
GN SEC5 OR YDR166C OR YD9489.01C OR YD8358.20C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 69-80.  
RX MEDLINE=97133278; PubMed=8978675;  
RA TerBush D.R., Maurice T., Roth D., Novick P.;  
RT "The Exocyst is a multiprotein complex required for exocytosis in  
RL Saccharomyces cerevisiae";  
RN EMBO J. 15:6483-6494(1996).  
RP [2]  
RP SEQUENCE OF 1-412 FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;  
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
P [3]  
P SEQUENCE OF 81-971 FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,  
RA Walsh S.V.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR EXOCYTOSIS.  
CC -!- SUBUNIT: SEC3, SEC5, SEC8, SEC10, SEC15 AND EXO70 ARE  
CC COMPONENT OF EXOCYST COMPLEX.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y08788; CAA70040.1; -  
DR EMBL; 247813; CAA87797.1; -  
DR EMBL; 230046; CAA90386.1; -  
DR SGD; S0002573; SEC5.  
KW Transport; Protein transport; Golgi stack; Coiled coil.  
FT DOMAIN 123 182 COILED COIL (POTENTIAL).  
SQ SEQUENCE 971 AA; 112121 MW; E41AC5A15CE3D1DF CRC64;  
  
Query Match 7.1%; Score 95; DB 1; Length 971;  
Best Local Similarity 21.6%; Pred. No. 6.8;  
Matches 64; Conservative 41; Mismatches 115; Indels 76; Gaps 13;  
  
QY 6 LSELIIDIGNTAFALFKDNQVNLFIKMTNLMRYDEVYSPFEENFDF-----NVN 58  
DB 133 LDRLSDIDQDSIHLKQLGVKNTFYKIRKNL-----DQIYKEFDEKTEKNOCDSPKN 188  
QY 59 KVFISVVPILNTEFNVI-FSPFKIKPLFGFD--LNYDIT-----ENPKYSDK 105  
DB 189 QINVES-----LNKKVDEVIRTTFKLPLMDNTQKILNYQATKFKIELNFKYPLPKSLK 244  
QY 106 FLGSDVFAANLVAIENYS-----FENVLVDLGTACTIFAVSRQDGILG 150  
DB 245 RCLTNDFNEFI--EYSGKLTLLRRFNQSSDASQSLVKRIWTQIENLLVTKDLIWN 301  
QY 151 GIINSGPLINFN-----SLLDNAYLKPKPISTPNLLE-----RTTSGVNSGL 195  
DB 302 SLINS-----NFNDIQPOETILSFKLLNLENF---INNQRSESGKNKNTSSSNENPI 354  
QY 196 FYQYKYLIEGVYRDIKMYKKKNLII-----TGGNADLILSLIEIFINIH 243  
DB 355 LRWMSIKMGFQNELNELSGHMSKIIHSORLIQNNTNODKSGCGVELSYLKLIN 410  
  
RESULT 13  
MVIN\_RICPR STANDARD; PRY; 507 AA.

AC 092CW4;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Virulence factor mvIN homolog.  
GN MVIN OR Rp590.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsieae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Madrid E.  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria";  
RL Nature 396:133-140(1998).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: BELONGS TO THE MVIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ235272; CAA15035.1; -  
DR InterPro; IPR004268; MVIN\_Like.  
DR Pfam; PF03023; MVIN; 1.  
KW Transmembrane; Complete proteome.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 92 112 POTENTIAL.  
FT TRANSMEM 132 152 POTENTIAL.  
FT TRANSMEM 157 177 POTENTIAL.  
FT TRANSMEM 185 205 POTENTIAL.  
FT TRANSMEM 245 265 POTENTIAL.  
FT TRANSMEM 268 288 POTENTIAL.  
FT TRANSMEM 310 330 POTENTIAL.  
FT TRANSMEM 351 371 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 405 425 POTENTIAL.  
FT TRANSMEM 438 458 POTENTIAL.  
FT TRANSMEM 472 492 POTENTIAL.  
SQ SEQUENCE 507 AA; 56059 MW; FFC56CCE2EC8EAC1 CRC64;  
  
Query Match 7.0%; Score 93.5; DB 1; Length 507;  
Best Local Similarity 22.0%; Pred. No. 4.1;  
Matches 63; Conservative 58; Mismatches 112; Indels 53; Gaps 14;  
  
QY 2 NKPLSELIDIGNTSTA-----FALFKDNQVNLFIKMTNLMRYDEVYSPFEENFDFN 56  
DB 220 NDPDVKKLLINMGPATISSGVQQLNLFISQSISSEIGAISILAYADRIYQPLS----I 275  
QY 57 VNKVFISVVPILNETFK-NVIFSPFKIK--PLFIFDNLVDLTENPKYKSDKFLGSDVF 113  
DB 275 IGTSFSTILLPEMSKVYKNDIVSAQKQNNAIRIGLLSLPATFG-----IILSHPI 329  
QY 114 ANLVAAIENYSFENVLVVDLGTACTIFAVSRQDGILGIIINSGPLINFNSLDNAYLK 173  
DB 330 TNII--YERGVTTPQDTTNTAEIASAPALGLPAFLAKILT--PIFYANGDTKPLKITL 385  
QY 174 FP--ISFPNNLL-----ERTTSGVNSGLFYQK-----YVIEGVYRDIK 213  
DB 386 FSIITINTNMILLMDLSLKHIGIAVGTSTAAVNLGLLYSYSTKOKHLIEAGI-----KL 440  
QY 214 YKKFNLIITGGNADLILSLIE---IEFIFNHLTVEGVRLIGNSI 256

441 FCAKILLCT--LMSIIIALKHVLYSEYLLIK-VSMIGSTI 483

DB

RESULT 14

Y468\_MYCGE

ID Y468\_MYCGE STANDARD; PRT; 1783 AA.

AC Q49460;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG468.

GN MG468.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI\_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RA MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium.";

RL Science 270:397-403(1995).

RN [2]

RP REVISIONS.

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 879-985 FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RA MEDLINE=94075230; PubMed=8253680;

RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;

RT "A survey of the Mycoplasma genitalium genome by using random sequencing.";

RL J. Bacteriol. 175:7918-7930(1993).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: SOME, TO MG064.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; U39728; AAC72488.1; -

DR EMBL; U01808; AAD12339.1; -

DR TIGR; MG468; -

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 16 36 POTENTIAL.

FT TRANSMEM 917 937 POTENTIAL.

FT TRANSMEM 967 987 POTENTIAL.

FT TRANSMEM 1010 1030 POTENTIAL.

FT TRANSMEM 1084 1104 POTENTIAL.

FT TRANSMEM 1660 1680 POTENTIAL.

FT TRANSMEM 1709 1729 POTENTIAL.

FT TRANSMEM 1730 1750 POTENTIAL.

FT TRANSMEM 1752 1772 POTENTIAL.

SQ SEQUENCE 1783 AA; 200168 MW; 87BD5AEAC2E374B CRC64;

Query Match 7.0%; Score 93.5; DB 1; Length 1783;

Best Local Similarity 22.1%; Pred. No. 18;

Matches 60; Conservative 36; Mismatches 81; Indels 95; Gaps 12;

17 SIAPALFKDQVNLFIKMTNLMRLRYDE-----YISFFEEFND-----FNVNKKV----- 60

838 SILFAVPAANOENYAFKSTDLKQHTDQDPVOFIANRLEGYLDVPRSDLAENVDKSFN 897

61 -----FISSVVVPIILNEFKVNFISFFFKIPKPIFGFDLNDVLTFFN--- 99

898 YLTARNYFPDLVQSYLAIVSTVIAIF-----LIILALYLILILIKFSIKKNQTEFSIIR 952

100 --PYKSKDKFLGSDVFNALVAIENYSFENVVVDLGTACTIFAVSRDGLGGLGINS GP 157

953 AGGFSTYKFIIVGMSVFAGIVAVS--SFLGVLPFAFL-----LEGQVKGILN--- 996

158 LINFNSLLDNAYLIKFKPPISTPNNLLERTTSGSVN----SGLFYQYKYLIEGV-YRDIKQ 212

997 -----RYWFIALPEN-----SPNLSFFGSGFTITTFVPEFISWIAFKQ 1034

213 MYKKFNLIITGGNADLILSLIEFIEFNIHL 244

1035 LFSKPVNVLIDQGN-----ETKFSVLLHL 1058

RESULT 15

ANFK\_AZOVI

ID ANFK\_AZOVI STANDARD; PRT; 462 AA.

AC P16267;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Nitrogenase iron-beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 beta subunit).

GN ANFK.

OS Azotobacter vinelandii.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Azotobacter.

OX NCBI\_TaxID=354;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=89123105; PubMed=2644222;

RA Joergers R.D., Jacobson M.R., Premakumar R., Wolfinger E.D., Bishop P.E.;

RT "Nucleotide sequence and mutational analysis of the structural genes (anfHDKG) for the second alternative nitrogenase from Azotobacter vinelandii.";

RL J. Bacteriol. 171:1075-1086(1989).

CC -!- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE IRON PROTEIN (COMPONENT 2) AND A COMPONENT 1 WHICH IS EITHER A MOLYBDENUM-IRON PROTEIN, A VANADIUM-IRON, OR AN IRON-IRON PROTEIN.

CC -!- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.

CC -!- COFACTOR: IRON-SULFUR.

CC -!- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO DELTA CHAINS.

CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFEN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; M23528; AAA82511.1; -

DR FIR; D32057; D32057.

DR HSP; P11347; LMIO.

DR InterPro; IPR000318; Nitrognase\_compl.

DR InterPro; IPR000510; Oxred\_nitrognasel.

DR Pfam; PF00148; Oxidored\_nitro; 1.

DR PROSITE; PS00090; NITROGENASE\_1\_2; 1.

DR PROSITE; PS00699; NITROGENASE\_1\_1; 1.

KW Oxidoreductase; Nitrogen fixation; Iron-sulfur.

SQ SEQUENCE 462 AA; 51179 MW; 2E4DE6267094E3CC CRC64;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.55623 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453a-11  
Perfect score: 1331  
Sequence: 1 MNKPLLSLLIDIGNTSIAF.....HLTVEGVRIILGNSIDFKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1331	100.0	262	2 F70165	conserved hypothet
2	289.5	21.8	273	2 E97293	probable transcrip
3	275.5	20.7	254	2 F83660	hypothetical prote
4	253.5	19.0	259	2 AF1464	conserved hypothet
5	248.5	18.7	259	2 AF1102	conserved hypothet
6	231.5	17.4	233	2 S66100	conserved hypothet
7	203	15.3	246	2 D72320	conserved hypothet
8	189.5	14.2	261	2 B87489	transcription acti
9	181.5	13.6	265	2 T36391	hypothetical prote
10	170	12.8	273	2 D71326	conserved hypothet
11	159	11.9	262	2 E75316	conserved hypothet
12	152.5	11.5	224	2 A99571	conserved hypothet
13	143	10.7	274	2 H86937	conserved hypothet
14	141.5	10.6	229	2 A70465	hypothetical prote
15	138	10.4	272	2 A70955	hypothetical prote
16	121.5	9.1	209	2 H81382	hypothetical prote
17	120.5	9.1	476	2 A28439	endonuclease Scl
18	118	8.9	276	2 A12292	hypothetical prote
19	118	8.9	625	2 A90127	hypothetical prote
20	116	8.7	2178	2 S55805	alpha-toxin - Clos
21	112.5	8.5	963	2 C90535	conserved hypothet
22	112	8.4	636	2 F90094	hypothetical prote
23	111.5	8.4	629	2 T28217	hypothetical prote
24	110.5	8.3	1003	2 H82883	hypothetical prote
25	110	8.3	300	1 S41759	ribosomal protein
26	109.5	8.2	2136	2 A05037	hypothetical prote
27	108	8.1	311	2 D97058	SAM-dependent meth
28	107.5	8.1	1613	2 D90129	hypothetical prote
29	107	8.0	590	2 G90127	hypothetical prote

30 103.5 7.8 1581 2 B71636 hypothetical prote  
31 103 7.7 436 2 B70157 hypothetical prote  
32 103 7.7 1465 2 A70199 hypothetical prote  
33 102 7.7 1024 1 RNZQBF DNA-directed RNA p  
34 101.5 7.6 594 2 A82913 hypothetical prote  
35 101.5 7.6 1119 2 T18491 hypothetical prote  
36 101.5 7.6 1802 2 G71616 hypothetical prote  
37 101 7.6 1175 2 F64489 hypothetical prote  
38 100.5 7.6 495 2 E70198 hypothetical prote  
39 100.5 7.6 512 2 F71915 hypothetical prote  
40 100.5 7.6 568 2 D90525 hypothetical prote  
41 100 7.5 355 2 H71730 hypothetical prote  
42 100 7.5 592 2 B81009 BirA protein/Bvg a  
43 99.5 7.5 410 2 S72277 translation elonga  
44 99.5 7.5 439 2 T28196 hypothetical prote  
45 99.5 7.5 516 2 S34525 hypothetical prote

## ALIGNMENTS

## RESULT 1

F70165  
conserved hypothetical protein BB0527 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: F70165  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; W  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; V  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, E  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: F70165  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <KLE>  
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:RAC66882.1; PID:g26  
A:Experimental source: strain B31

Query Match 100.0% Score 1331; DB 2; Length 262;  
Best Local Similarity 100.0%; Pred. No. 5.6e-94;  
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFLTKMTNMLRYDEVYSFEEFDFNNKV 60  
Db 1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFLTKMTNMLRYDEVYSFEEFDFNNKV 60  
QY 61 FISSVVPILNETFKNVIFFPKIKPLFIFGDLNYDLTFNPKSKDKLLGSDVFANLVAAL 120  
Db 61 FISSVVPILNETFKNVIFFPKIKPLFIFGDLNYDLTFNPKSKDKLLGSDVFANLVAAL 120  
QY 121 ENSPENVLVDLGTACTIFAVSRDGLGGIINSGLINFLNLSLLNAYLIKFPSTPN 180  
Db 121 ENSPENVLVDLGTACTIFAVSRDGLGGIINSGLINFLNLSLLNAYLIKFPSTPN 180  
QY 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIF 240  
Db 181 NLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEIEFIF 240  
QY 241 NIHLTVEGVRIILGNSIDFKFVN 262  
Db 241 NIHLTVEGVRIILGNSIDFKFVN 262

## RESULT 2

E97293  
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clos  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: E97293  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;

;; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97293  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3200  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 21.8%; Score 289.5; DB 2; Length 273;  
Best Local Similarity 29.6%; Pred. No. 7.5e-15;  
Matches 77; Conservative 62; Mismatches 96; Indels 25; Gaps 8;

QY 9 LIIDIGNTSIAFALFKDQNVNLFIRKMTNLMRLRYDE---VYSFFEEN-FDEN-VNKVFI 62  
... 14 LVIDGNTNIVIGIYNDTKLTAEWRLSTDVLSRSDYEGIOVMNLFQODKLDPTLVGVI 73

QY 63 SSVVPILNETFNKVFISFEKIKPLFI-----GFDLNDLTNPYKSKDKFLGSDVFANL 116  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 74 SSVVPNIMYSLEHMIRKYEKINPLVVGPIKGTGINKYD---NPKE-----VGADRIVNA 125  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 117 VAAIENYSPENVLVVDLGTACTIFAVSRQDGLGGIINSGLPINFNSLDDNAYLIKPKFI 176  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 126 VAAHEIYK-RSLIIIDFGTATTFCVAVRENGDYLGAICPGIKVSSSEALFEKAAPRVEL 184  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 177 SPNNLLERTTSGSVNSGLFYQY---KYLIEGVYRDIKQMKKFNLIITGGNADLIIS 232  
Db 185 IKPAYAICKNTISSQSGIVYIGQVRIYERMKEEQEKEPVLVATGGLAKLISE 244  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 233 LIEIEFIFNIHLTVGVRIL 252  
Db 245 EAKNDVINPFTLGLRII 264  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
RESULT 3  
F33660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F33660  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
Accession: F33660  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <STO>  
A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH0086  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 20.7%; Score 275.5; DB 2; Length 254;  
Best Local Similarity 28.1%; Pred. No. 8e-14;  
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;

QY 9 LIIDIGNTSIAFALFKDQNVNLFIRKMTNLMRLRYDEVYSFFEENFDF-----NVNKVFI 62  
3 LVIDGNTNIVIGYQDETLLVHHWRLATSRQKTEDEYAMTVRSLEFDHAGLQFQDIDGIVI 62  
Db 63 SSVVPILNETFNKVFISFEKIKPLFI-----GFDLNDLTNPYKSKDKFLGSDVFANL 116  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 63 SSVVPPMFMFSLEQMKCKYFHTVPTMIIGPGIKTGLNLIKVD---NPKE-----VGADRIVNA 114  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 117 VAAIENYSPENVLVVDLGTACTIFAVSRQDGLGGIINSGLPINFNSLDDNAYLIKPKFI 176  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
RESULT 4  
AF1464  
conserved hypothetical protein lln0253 [imported] - Listeria innocua (strain Clip112  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A:Title: Comparative genomics of Listeria species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:AL502022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lln0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 19.0%; Score 253.5; DB 2; Length 259;  
Best Local Similarity 28.1%; Pred. No. 3.8e-12;  
Matches 71; Conservative 53; Mismatches 114; Indels 15; Gaps 5;

QY 9 LIIDIGNTSIAFALFKDQNVNLFIRKMTNLMRLRYDE---VYSFFE--ENFDNFVNKVF 62  
1:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 3 LVIDGNTNCTGVYKEQKLLRHRWMTDRHRTSDELGMTVLFNFFSVANLTPSDIOGIII 62  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 63 SSVVPILNETFNKVFISFEKIKPLFIETGFDLNDLTFT---NPYKSKDKFLGSDVFANVAA 119  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 63 SSVVPPIMHAMETCMCYRNIPLVVGPIKGTGLNLIKVDNPRE-----IGSDRIVNAAA 117  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 120 IENYSPENVLVVDLGTACTIFAVSRQDGLGGIINSGLPINFNSLDDNAYLIKPKFTSTP 179  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 118 SEEIG-TPVIVVDGFTATTCYIDEAGVYOGGAIAPGIMISTEALYNRAAKLPRVDIAES 176  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 180 NNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMKKFNLIITGGNADLIISLIEIFI 239  
1:||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
Db 177 SGIICKSTVASQAGIFYGIGCEGIIAEMKKQSNTSPVVATGGLARMITEKSSAVDI 236  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
QY 240 FNIHLTVGVRIL 252  
Db 237 LDPFLTLKGLLEL 249  
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
RESULT 5  
AF1102  
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EK  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1102  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <GLA>

A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: imo221

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 18.7%; Score 248.5; DB 2; Length 259;  
Best Local Similarity 27.7%; Pred. No. 9.2e-12;  
Matches 70; Conservative 53; Mismatches 115; Indels 15; Gaps 5;

Qy 9 LIIDIGNTSIAFALEKDNQVNLFIKMTNLMRLRYDE---VYSFFE--ENPDFNVNKFVI 62

Db 3 LVIDVGNNTCVGYEQKLLKHWRTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

y 63 SSVVPILNETFKNVIFSEFFKPLFIQFDNLNYDTF---NPYKSKDFLLGSDVFANLVA 119

Db 63 SSVVPPIMHAMETMCMVRYFNIRPLIVGPGIKTGLNLKVDNPRE-----IGSDRVNVA 117

Qy 120 IENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLIKKEPISTP 179

Db 118 SEEGV-TPVIIVDEGTATTCYIDESGYGGATAGIMISTEALYNAAKLPVDAES 176

Qy 180 NNLLERTTSGVNSGLFYQYKYLIEGVYRDIKOMYKKFNLIITCGNADLILSLIEIFI 239

Db 177 SQIIIGKSTVSMQAGIFYGVGOCEGIIAEMKQSNASPVVATGGLARMITERKSSAVDI 236

Qy 240 FNHLTVEGVRIL 252

Db 237 LDPLFLTKGLELL 249

#### RESULT 6

S66100

conserved hypothetical protein yacB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000

C:Accession: S66100; E69740

R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994

A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom

A:Reference number: S65967; MUID:96051385; PMID:7584024

A:Accession: S66100

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <OGA>

A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BA005305.1; PID:d1005847; PID:g467459

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emerson, P.R.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koetter, P.; Koningsberg, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, J.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: E69740

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <KUN>

A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182003;

A:Experimental source: strain 168

C:Genetics:

A:Gene: yacB

A:Start codon: TTG

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 17.4%; Score 231.5; DB 2; Length 233;  
Best Local Similarity 26.8%; Pred. No. 1.6e-10;  
Matches 59; Conservative 48; Mismatches 92; Indels 21; Gaps 5;

Qy 9 LIIDIGNTSIAFALEKDNQVNLFIKMTNLMRLRYDEVYSFFENPDFN-----VNKVF 62

Db 3 LVIDVGNNTVGLVTHDGKLEHYHRTSRHRTDEFGMLLSLFDHSGLMFEQDIIII 62

Qy 63 SSVVPILNETFKNVIFSEFFKPLFI-----GFDLNYDLTFNPKSKDFLLGSDVFANL 116

Db 63 SSVVPPIMFALERMTKVFHIEPQIVGPGMKTLNLIKVD---NPKR-----VGADRIVNA 114

Qy 117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLIKKEFPI 176

Db 115 VAAIHLYG-NPLIVVDFGTATTCYIDENKQTMGGAIAPGITISTEALYSRAAKLPRIE 173

Qy 177 STPNLLERTTSGVNSGLFYQYKYLIEGVYRDIKOMYKK 216

Db 174 TRPDNIIGNKTVSAMQSGLIFGVGVQVEGIVAKMKWQAKQ 213

#### RESULT 7

D72320

conserved hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: D72320

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72320

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-246 <ARN>

A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g49

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0883

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 15.3%; Score 203; DB 2; Length 246;  
Best Local Similarity 24.6%; Pred. No. 2.5e-08;  
Matches 60; Conservative 62; Mismatches 112; Indels 10; Gaps 7;

Qy 9 LIIDIGNTSIAFALEKDNQVNLFIKMTNLMRLRYDEVYSFFEE---NFDENVKNVFISSV 65

Db 3 LLVDVGNTHSVFSTEDGKTFRRWLSTGVTFQTEDELFSLHPLLDGDMREIKGIVASV 62

Qy 66 VPILNETFKNVIFSEFFKPLFIQFDNLVDFNFPYKSKDFLLGSDVFANLVAIENYSF 125

Db 63 VPTQNTVIERSQKVFHISPLVWAKK-NGCVKWNKNPSE--VGADRANVAVFAVKEYG- 118

Qy 126 ENLVVDLGTACTIFAVSRQDGLIGLINSGLPLNFNSLLDNAYLIKKEPISTPNLLER 185

Db 119 KNGIIIDGTATTVDLVV-NGSYEGGAILPGFFMMVHSLFRGTAKPLPLVEKVPADPEVVGK 177

Qy 186 TTSSVNSGLFYQYKYLIEGVYRDIKOMYKKFNLIITCGNADLILSLIEIEIFNIHLT 245

Db 178 DTEENIRLGVVNSVYALEGIIGRIKEYV-GDLPVVLVTGGQSKIVKMDIKHE-IFDEDLT 235

Qy 246 VEGV 249

Db 236 IKGV 239





Db 111 ANAYVVL--HKSKNAIFISLGTAVITQIKNSSIEGSIYPCIQSPKFNFWAKIESN 168  
| | : : : : : ||||| : : | | | | : : : :  
Qy 169 YLKPKFPISPNNLLERTTSGSVNSGLFYQKYKLIEGYRDIKOMYKKFNLITGG 225  
| | : : : : : ||||| : : | | | | : : : :  
Db 169 FNIKIP-----PILGKTTLTLESLSLVGRGSVFLLKGVIDEI----DKTSDFITGG 216  
| | : : : : : ||||| : : | | | | : : : :

RESULT 13

H86937 conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
C/Species: Mycobacterium leprae  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C/Accession: H86937  
R/Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holr  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.  
A/Title: Massive gene decay in the leprosy bacillus.  
A/Reference number: A86909; MUID:21128732; PMID:11234002  
A/Accession: H86937  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-274 <STO>  
A/Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147  
C/Genetics:  
A/Gene: ML0232  
C/Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 10.7% Score 143; DB 2; Length 274;  
Best Local Similarity 19.2%; Pred. No. 0.001;  
Matches 51; Conservative 72; Mismatches 111; Indels 32; Gaps 8;

Qy 9 LIIDIGNTSTAFALEKDNQVNLF-----KMTNLMRYDEYVSFFE-----ENFDENVVKY 60  
| | | | : : : : : ||||| : : | | : : : :  
Db 3 LAIDVRNTHRWVLLSGSKSHAKVVQQWRINTSEVTADALAIJDIGLDSDERLAGAA 62  
| | | | : : : : : ||||| : : | | : : : :  
Qy 61 FISWVPILNETKFNVIFFFKIKPLFI-----GPDNLYDLTFNPYKSKDKFLIGSDVF 113  
| | | | : : : : : ||||| : : | | : : : :  
Db 63 ALSTVPSVLHEV-RIMLDQWSPVPHVLPBGVFTGPLVD---NPKE-----VGADRI 113  
| | | | : : : : : ||||| : : | | : : : :  
Qy 114 ANLVAATNTSFENVLDGTACTIFAVSRQDGIIGIINSPLINFSLLDNNAYLK 173  
| | | | : : : : : ||||| : : | | : : : :  
Db 114 VNCLAAPHKFG-QAAIVDFGSSICVDWSAKGEFLGAGIAPGVQVSSDAARAASALRR 172  
| | | | : : : : : ||||| : : | | : : : :  
Qy 174 FPISTPNLLERTTSGSVNSGLFYQKYKLIEGVYRDIKOMYKK-----KENLIITGN 226  
| | | | : : : : : ||||| : : | | : : : :  
Db 173 VELAPRSVVGKNTECMQAGVFEFGAGLDGLVGRMQDVVEEFGSLDGNRVAVVATGHT 232  
| | | | : : : : : ||||| : : | | : : : :  
Qy 227 ADILSLIETEFINHLTVEGVRIL 252  
| | | | : : : : : ||||| : : | | : : : :  
Db 233 APILLPELHVHDHRLTLHLGLRV 258  
| | | | : : : : : ||||| : : | | : : : :

RESULT 14

E70465 hypothetical protein aq\_1924 - Aquifex aeolicus  
C/Species: Aquifex aeolicus  
C/Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C/Accession: E70465  
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.  
V.

Nature 392, 353-358, 1998  
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A/Reference number: A70300; MUID:98196666; PMID:9537320  
A/Accession: E70465  
A/Status: preliminary;  
A/Molecule type: DNA  
A/Residues: 1-229 <AQF>  
A/Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE0  
A/Experimental source: strain VF5  
C/Genetics:

search completed: June 1962  
Job time : 10.5562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.899 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-11  
Perfect score: 1331  
Sequence: 1 MNKPLLELLIIGNTSIAF.....HLTVEGVRILGNSIDPKFVN 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1331	100.0	262	AAU91158	Borrelia burgdorferi
2	282	21.2	256	AAU91175	Pantothenate kinas
3	276.5	20.8	262	AAU91170	Pantothenate kinas
4	275.5	20.7	254	AAU91171	Pantothenate kinas
5	271.5	20.4	258	AAU01243	B. subtilis novel
6	271.5	20.4	258	AAU91149	Bacillus subtilis
7	266.5	20.0	255	AAU91154	Geobacter sulfurre
8	262	19.7	250	AAU91150	Clostridium acetob
9	256.5	19.3	258	AAU91172	Pantothenate kinas
10	248.5	18.7	259	ABB47661	Listeria monocytog

11	231.5	17.4	233	AAU91163	Pantothenate kinas
12	203	15.3	246	AAU91156	Thermotoga maritim
13	197.5	14.8	258	AAU91153	Rhodobacter capsul
14	190	14.3	219	AAU91176	Pantothenate kinas
15	189.5	14.2	260	AAU91173	Pantothenate kinas
16	181.5	13.6	265	AAU91151	Streptomyces coeli
17	170	12.8	273	AAU91157	Treponema pallidum
18	166.5	12.5	257	AAU91174	Pantothenate kinas
19	159	11.9	262	AAU91155	Deinococcus radiop
20	141.5	10.6	229	AAU91159	Mycofex aeolicus p
21	138	10.4	272	AAU91152	Mycofex aeolicus p
22	138	10.4	272	AAU91152	Mycofex aeolicus p
23	136	10.2	241	AAU91179	Mycofex aeolicus p
24	135	10.1	244	AAU91168	Pantothenate kinas
25	121.5	9.1	209	AAU91165	Pantothenate kinas
26	121.5	9.1	476	AAU91165	Pantothenate kinas
27	121.5	9.1	503	AAU91165	Pantothenate kinas
28	111	8.3	212	AAU91177	50 kD subunit of S
29	104	7.8	471	AAU91177	Pantothenate kinas
30	101.5	7.6	1802	AAU91177	Amino acid sequenc
31	100	7.5	389	AAU91177	Plasmodium falcipa
32	100	7.5	592	AAU91177	Neisseria meningit
33	100	7.5	592	AAU91177	Neisseria meningit
34	100	7.5	592	AAU91177	Neisseria meningit
35	99.5	7.5	410	AAU91177	Pantothenate kinas
36	99.5	7.5	410	AAU91177	Plasmodium falcipa
37	98.5	7.4	468	AAU91177	Plasmodium falcipa
38	97.5	7.3	1188	AAU91177	AmEPV serine threo
39	96.5	7.3	439	AAU91177	Plasmodium falcipa
40	96	7.2	435	AAU91177	Streptococcus poly
41	96	7.2	435	AAU91177	Neisseria meningit
42	96	7.2	592	AAU91177	Neisseria meningit
43	96	7.2	592	AAU91177	Neisseria meningit
44	95.5	7.2	810	AAU91177	Pantothenate kinas
45	95	7.1	451	AAU91177	Fungal ZBC protein
					WAR toxin from B.

## ALIGNMENTS

RESULT 1  
AAU91158  
ID AAU91158 standard; Protein; 262 AA.  
XX AC AAU91158;  
XX AC

05-JUN-2002 (first entry)

DE Borrelia burgdorferi pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS

XX Borrelia burgdorferi.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

XX PS  
XX CC  
XX CC

Claim 6; Page 76; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX SQ Sequence 262 AA;

Query Match 100.0%; Score 1331; DB 23; Length 262;

Best Local Similarity 100.0%; Pred. No. 3.1e-127; Mismatches 0; Indels 0; Gaps 0;

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDENVKV 60  
1 MNKPLLSLLIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDENVKV 60  
61 FISSVVPILNETKFNVIFFSKIKPLFLFGDLNLDLTNPKSKDFLLGSDVFANLVAI 120  
61 FISSVVPILNETKFNVIFFSKIKPLFLFGDLNLDLTNPKSKDFLLGSDVFANLVAI 120  
121 ENYSFENVLVVDLTACTIFAVSRDGLGGLINSGLPLNLSLDNAYLIKPFISTPN 180  
121 ENYSFENVLVVDLTACTIFAVSRDGLGGLINSGLPLNLSLDNAYLIKPFISTPN 180  
181 NLLERTSGVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGNADLIILSIEIEIF 240  
181 NLLERTSGVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGNADLIILSIEIEIF 240  
241 NIHLTVEGVRLGNSIDPKFVN 262  
241 NIHLTVEGVRLGNSIDPKFVN 262

RESULT 2

AAU91175  
ID AAU91175 standard; Protein; 256 AA.

XX AC AAU91175;

XX DT 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #13.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX OS Clostridium difficile.

XX FN WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX DR N-PSDB; ABK54196.

XX PT Identifying potential antibiotic or antimicrobial agent, comprises

PT XX  
PT PS  
XX CC  
XX CC

contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein

Claim 6; Page 105; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

XX SQ Sequence 256 AA;

Query Match 21.2%; Score 282; DB 23; Length 256;

Best Local Similarity 31.3%; Pred. No. 2e-20;

Matches 82; Conservative 54; Mismatches 94; Indels 32; Gaps 10;

QY 9 LIIDIGNTSIAFALFKDNQVNLFIKMTNLMRLRYDEVYSFFEENFDENVKVFI 62  
DB 3 LVFDVGNMVLGKIKPLFLFGDLNLDLTNPKSKDFLLGSDVFANLVAI 62  
QY 63 SSVVVPILNETKFNVIFFSKIKPLFLFGDLNLDLTNPKSKDFLLGSDVFANL 116  
DB 63 SSVVVPILNETKFNVIFFSKIKPLFLFGDLNLDLTNPKSKDFLLGSDVFANL 114  
QY 117 VAAIENYFENVLVVDLTACTIFAVSRDGLGGLINSGLPLNLSLDNAYLIKPF 176  
DB 115 VAGIEKYGAPSL-VDFGTATTFCATSEKGYLGTTAFGKISSALFQSASKLPRVEL 173  
QY 177 STPNLLERTSGVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGNADLIIL 231  
DB 174 AKPGMTICKSTVSAMQSGIIVGVGLVD---KIISIMKELNCCDDVKVIATGGLAKLIA 229  
QY 232 SLIE-IEFIFNIHLTVEGVRL 252  
DB 230 SETKSIDYVDGF-LTLEGLRII 250

RESULT 3

AAU91170

ID AAU91170 standard; Protein; 262 AA.

XX AC AAU91170;

XX DT 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #8.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX OS Bacillus anthracis.

XX FN WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX DR N-PSDB; ABK54191.

```

XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 8; Page 98-99; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 262 AA;
Query Match 20.8%; Score 276.5; DB 23; Length 262;
Best Local Similarity 27.8%; Pred. No. 7.6e-20;
Matches 71; Conservative 63; Mismatches 100; Indels 21; Gaps 7;
QY 10 IIDIGNSTIAFAFKDQVNLFIKMKTNLMRLRYDE---VYSFFE-ENPDF-NVNVKVFIS 63
Db 4 VLDVGNTNAVGVFEEGELRQHRMETDRHKTDEYGMVYKQLLEHGLSFEDVKGIIVS 63
QY 64 SVVPILNETFKNVIFSFKKIKPLFI-----GFDLVNLTFFNPKSKDKFLGSDVFANLV 117
Db 64 SVVPTMFALERMCERYFKIKPLVGVGPKTGKTNKIYE---NPRE-----VGADRVNAV 115
QY 118 AAIENYSFENLVLDLTACTIFAVSRQDGIILGIIINSGLINFNLSLDNAYLIKRPPIIS 177
Db 116 AGIHLVG-SPLIIVDFGTATTCYINEERHYMGVITPGIMISAEALYSRAAKLPRIET 174
QY 178 TPNNLLERTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKFNLIITGGNADLILSLIEIE 237
Db 175 KPSSVGVKNTVSAMQSGIILYGVGVQEGIVKRMKEAKQPKVIATGGLAKLISESNVI 234
QY 238 FIFNIHLTVGVRIL 252
Db 235 DVVDPELTGLKGLYL 249
RESULT 4
AAU91171
AAU91171 standard; Protein; 254 AA.
AC AAU91171;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #9.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus halodurans.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
XX
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX

```

```

DR WPI; 2002-269358/31.
DR N-PSDB; ABK34192.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 10; Page 100; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 254 AA;
Query Match 20.7%; Score 275.5; DB 23; Length 254;
Best Local Similarity 28.1%; Pred. No. 9.2e-20;
Matches 72; Conservative 58; Mismatches 105; Indels 21; Gaps 5;
QY 9 LIIDIGNSTIAFAFKDQVNLFIKMKTNLMRLRYDEYVSFFENPDF-----NVNVKVF 62
Db 3 LVIDVGNTNVGVYQDETIVHWRLATSRQKTEDEVAMTVRSFLFDHAGLQFQDIDGIVI 62
QY 63 SSVVPILNETFKNVIFSFKKIKPLFI-----GFDLVNLTFFNPKSKDKFLGSDVFANLV 116
Db 63 SSVVPMFMFLEQCKKRYFHTPMIIGPKTGLNIKYD---NPKR-----VGADRVINA 114
QY 117 VAALENYSFENLVLDLTACTIFAVSRQDGIILGIIINSGLINFNLSLDNAYLIKRPPI 176
Db 115 VAALYGY-P-ALVVDFTATTTCYLINEKKQYAGGVAPGIMISTEALYHRASKLPRIE 173
QY 177 TPNNLLERTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKFNLIITGGNADLILSLIEI 236
Db 174 AKPKQVVGTTIDSMQSGIFVGVYQVGVKRMKAQAESEPKVIATGGLAKLIGTSET 233
QY 237 EFIFNIHLTVGVRIL 252
Db 234 IDVIDSELTGLQLI 249
RESULT 5
AAU01243
AAU01243 standard; Protein; 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coax.
XX
KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
XX
PR 07-JUN-2000; 2000US-0210072.
XX
PR 28-JUL-2000; 2000US-0221836.
XX
PR 24-AUG-2000; 2000US-0227860.
XX

```

PA (OMNI-) OMNIGENE BIOPRODUCTS.  
 PI Yocum RR, Patterson TA, Hermann T, Pero JG;  
 XX WPI; 2001-218644/22.  
 DR N-PSDB; AAS00984.  
 XX New recombinant microorganism which overexpress a Bacillus subtilis  
 PT pantothenate biosynthetic enzyme, useful for the high yield production  
 PT of panto-compounds such as pantothenate and pantoate.  
 XX Example 14; Fig 23; 292pp; English.  
 PS The sequence represents a novel B. subtilis pantothenate kinase (encoded  
 CC by gene cox), an enzyme of the pantothenate biosynthetic pathway.  
 CC Pantothenate, also known as vitamin B5, is used as a nutritional  
 CC supplement in mammals and humans. The invention concerns methods of  
 CC producing recombinant microorganisms overexpressing at least one Bacillus  
 CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
 CC of producing them are useful for producing a panto-compound such as  
 CC pantothenate or pantoate, which is a nutritional requirement for  
 CC livestock and humans. The methods are also useful for the identification  
 CC of pantothenate kinase modulators. Panto-compounds are produced at a  
 CC significantly higher yield than prior art methods and can be produced  
 CC independent of the need to feed precursors which decreases expense.  
 XX Sequence 258 AA;  
 SQ

Query Match 20.4%; Score 271.5; DB 22; Length 258;  
 Best Local Similarity 27.3%; Pred. No. 2.4e-19;  
 Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;  
 QY 9 LIIDIGNTSIAPALFKDNQVNLFIKMTNLMRLYDEVYSPFEENFDN-----VNKVF 62  
 Db 3 LVIDGNTVNLGVYHDGLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGIII 62  
 QY 63 SSVVPIINLNETFNKVFSPFKIKPLFI-----GFDLNYDLTFNPKSKDFLGSDFVFN 116  
 Db 63 SSVVPIINLNETFNKVFSPFKIKPLFI-----GFDLNYDLTFNPKSKDFLGSDFVFN 116  
 QY 117 VAAIENYSPENLVVDLGTACTTFAVSRQDGLGILGINSGLPLNFSLDNAYLIKKEPI 176  
 Db 115 VAAIHLYG-NPLIVDFGTATTTCYIDENKQYMGGAIAPIGISTEALYSRAAKLPRIE 173  
 QY 177 STPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIE 236  
 Db 174 TRPDNIIGNTVSAMQSGILFGYGVQGEIVKRMKQAKODLKVATGGLAPLIANESDC 233  
 QY 237 EFIFNIHLTVEGVRI 252  
 Db 234 IDIVDPFLTKGLELI 249

RESULT 6  
 AAU91149  
 ID AAU91149 standard; Protein; 258 AA.  
 XX AAU91149;  
 AC AAU91149;  
 DT 05-JUN-2002 (first entry)  
 DE Bacillus subtilis pantothenate kinase Coax.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Bacillus subtilis.  
 OS WO200216601-A2.  
 \*PN 28-FEB-2002.  
 PD 24-AUG-2001; 2001WO-US26531.  
 XX PF

XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 PA Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 DR N-PSDB; ABK54168.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 67-68; 128pp; English.  
 PS The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 258 AA;  
 SQ

Query Match 20.4%; Score 271.5; DB 23; Length 258;  
 Best Local Similarity 27.3%; Pred. No. 2.4e-19;  
 Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;  
 QY 9 LIIDIGNTSIAPALFKDNQVNLFIKMTNLMRLYDEVYSPFEENFDN-----VNKVF 62  
 Db 3 LVIDGNTVNLGVYHDGLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGIII 62  
 QY 63 SSVVPIINLNETFNKVFSPFKIKPLFI-----GFDLNYDLTFNPKSKDFLGSDFVFN 116  
 Db 63 SSVVPIINLNETFNKVFSPFKIKPLFI-----GFDLNYDLTFNPKSKDFLGSDFVFN 116  
 QY 117 VAAIENYSPENLVVDLGTACTTFAVSRQDGLGILGINSGLPLNFSLDNAYLIKKEPI 176  
 Db 115 VAAIHLYG-NPLIVDFGTATTTCYIDENKQYMGGAIAPIGISTEALYSRAAKLPRIE 173  
 QY 177 STPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIE 236  
 Db 174 TRPDNIIGNTVSAMQSGILFGYGVQGEIVKRMKQAKODLKVATGGLAPLIANESDC 233  
 QY 237 EFIFNIHLTVEGVRI 252  
 Db 234 IDIVDPFLTKGLELI 249

RESULT 7  
 AAU91154  
 ID AAU91154 standard; Protein; 255 AA.  
 XX AAU91154;  
 AC AAU91154;  
 DT 05-JUN-2002 (first entry)  
 DE Geobacter sulfurreducens pantothenate kinase Coax.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Geobacter sulfurreducens.  
 OS WO200216601-A2.  
 \*PN 28-FEB-2002.  
 PD 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 72-73; 128pp; English.  
 PS The invention describes assays for identifying a (potential) antibiotic  
 XX comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 255 AA;  
 SQ Query Match 20.0%; Score 266.5; DB 23; Length 255;  
 Best Local Similarity 29.3%; Pred. No. 7.7e-19;  
 Matches 75; Conservative 48; Mismatches 112; Indels 21; Gaps 5;  
 QY 9 LIIDIGNTAFALFKDQVNLFIKMTNLMRYDEVYSFFEEFDF-----NVNKFVI 62  
 DB 3 LVLDVGNNTVLGIYDGERLVRWVSTDKARTDEVGILNELFRLAGLGLQIRAVII 62  
 QY 63 SSVVPILNETFKNVIFSFYKIKPLFI-----GFDLNYDLTFNPKSKDFLLGSDVFANL 116  
 DB 63 SSVVPLTGLVRLSLGFCMRPLVVGPGIKTGMPIQYD---NPKE-----VGADRVNA 114  
 QY 117 VAAIENYSPENLVLDLGTACTIFAVSRQDGIILGGINSPILNFSLLDNAYLIKPFPI 176  
 DB 115 VAGYEYR-TSLIIVDFGTATTFDYNRKGECYCGAATAPGLVISTEALFORASKLPRVDI 173  
 177 STPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLILSLIEI 236  
 DB 174 IRPSAIIARTVNSMAGIYGVGLVDEIVTRKAEKSDAPRVATGGGLASLIAPESKT 233  
 QY 237 EFTFNHLTVGVRIL 252  
 DB 234 IEAVEEYLTLEGRIL 249  
 RESULT 8  
 AAU91150  
 ID AAU91150 standard; Protein; 250 AA.  
 XX  
 AC AAU91150;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX Clostridium acetobutylicum pantothenate kinase Coax.  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Clostridium acetobutylicum.  
 XX WO200216601-A2.  
 XX 24-AUG-2001; 2001WO-US26531.

PD 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 PF 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 68-69; 128pp; English.  
 PS The invention describes assays for identifying a (potential) antibiotic  
 XX comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 250 AA;  
 SQ Query Match 19.7%; Score 262; DB 23; Length 250;  
 Best Local Similarity 30.1%; Pred. No. 2.1e-18;  
 Matches 72; Conservative 54; Mismatches 87; Indels 26; Gaps 9;  
 QY 9 LIIDIGNTAFALFKDQVNLFIKMTNLMRYDE-----VYSFEEF-FDFN-VNKFVI 62  
 DB 20 LVLDVGNNTVLGIYDNTKLTAEWRLSTDVLSRDADEYGIQWNNLFGQDKLDTLVEGVII 79  
 QY 63 SSVVPILNETFKNVIFSFYKIKPLFI-----GFDLNYDLTFNPKSKDFLLGSDVFANL 116  
 DB 80 SSVVPNIMYSLEHMIRKVFKNPLVVGPGIKTGNIKYD---NPKE-----VGADRVNA 131  
 QY 117 VAAIENYSPENLVLDLGTACTIFAVSRQDGIILGGINSPILNFSLLDNAYLIKPFPI 176  
 DB 132 VAAHEIYK-RSLIIVDFGTATTFCAVRENGDYLGAICPGIKVSSEALFEKAAKLPRVEL 190  
 QY 177 STPNLLERTSGSVNSGLFYQYKYLIEGVYRDIKQYKKFNLIITGGNADLI 230  
 DB 191 IKPAAICKNTISSIQSGIVTRYLRQVKYLFKLEKLENLPDGRRTTSVLATGGLAKLI 249  
 RESULT 9  
 AAU91172  
 ID AAU91172 standard; Protein; 258 AA.  
 XX  
 AC AAU91172;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #10.  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Bacillus stearothermophilus.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.

```
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX PI WPI; 2002-269358/31.
XX DR N-PSDB; ABR54193.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein
XX PS Claim 10; Page 101-102; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 258 AA;
Query Match 19.3%; Score 256.5; DB 23; Length 258;
Best Local Similarity 26.3%; Pred. No. 8.1e-18;
Matches 67; Conservative 59; Mismatches 108; Indels 21; Gaps 5;
QY 10 IIDIGNTSIAFALFKDQVNLFIKMTNLMRYDEYISFFENFDF-----NVNKFVTS 63
DB 4 VLDVGNNTVLGVYDGDGLKHHWRIETSRKTEDEYGMKALLNHHVGLQFSDIRGIITS 63
QY 64 SVVPILNETFNKVFISFFKIKPLFI-----GFDLNYDLTFNPKYKSKFLLGSDVFANLV 117
DB 64 SVVPIPMFALERMCLYKFIKPLIYGVGQVIGVSRKAKSKIPPKVIATGGLAPLASESDII 115
QY 118 AAIENYSFENVLVLDGTACTIFAVSRQDGIILGGLIINSGLPINFNSLDNAYLIKFKPLTS 177
DB 116 AGIHLVG-SPLIIVDFGTATTTCYINHHQYMGGAIAFGIMISTEALFARAAKLPRIEIA 174
QY 178 TPNLLERTTSGVNSGLFYQYKYLIEGYRDIKQYKKNLITGGNADILSLIETIE 237
DB 175 RPDDIIGKNTVSAMAGILYGVQVIGVSRKAKSKIPPKVIATGGLAPLASESDII 234
DB 238 FIFNHLTVGVRIL 252
DB 235 DVVDPFLATLGLKL 249
RESULT 10
AB47661
ID AB47661 standard; Protein; 259 AA.
XX AC AB47661;
XX DT 05-FEB-2002 (first entry)
XX DE Listeria monocytogenes protein #365.
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX KW vitamin B12; bacterial infection; disease.
XX OS Listeria monocytogenes.
XX PN WO200177335-A2.
XX PD 18-OCT-2001.
```

```
XX PR 11-APR-2001; 2001WO-FR01118.
XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP ) INST PASTEUR.
XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX PI Dussurget O, Chetouni F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
XX PI Perez-Diaz J, Baquero E, Garcia Del Portillo F, Gomez-Lopez N;
XX PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
XX PI Rose M, Voss H;
XX DR WPI; 2002-010914/01.
XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX PT and prevention of Listeria and related bacterial infections, and
XX PT related polypeptides
XX PS Claim 6; SEQ ID NO 366; 192pp; French.
XX CC The present invention relates to the genome sequence of Listeria
XX CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
XX CC it are useful for selecting probes and primers for detecting genes in L.
XX CC monocytogenes and related organisms, and for studying genetic
XX CC polymorphisms and other genomes. The present sequence is a protein
XX CC encoded by the genome sequence of the present invention. Proteins
XX CC expressed from the genome sequence are useful for raising specific
XX CC antibodies, identification of L. monocytogenes and related organisms,
XX CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
XX CC B12. The genome sequence and proteins encoded by it are also useful for
XX CC selecting compounds that regulate gene expression and cell replication
XX CC and modulate L. monocytogenes-related diseases. In addition, the genome
XX CC sequence and proteins encoded by it are useful in pharmaceutical and
XX CC vaccine compositions for the treatment or prevention of infections by L.
XX CC monocytogenes and related organisms.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 259 AA;
Query Match 18.7%; Score 248.5; DB 23; Length 259;
Best Local Similarity 27.7%; Pred. No. 5.3e-17;
Matches 70; Conservative 53; Mismatches 115; Indels 15; Gaps 5;
QY 9 LIIDIGNTSIAFALFKDQVNLFIKMTNLMRYDE-----VYSFFE--ENFDENVKVF 62
DB 3 LVIDVGNNTCTVGVYEKQKLLKHHWRTTDRHRTSDELGMTVLNFFSYANLTPSDIOGIII 62
QY 63 SSVVPILNETFNKVFISFFKIKPLFIQFDLNYDLTF---NPYKSKDFLLGSDVFANLVAA 119
DB 63 SSVVPIMHAMETCMVRYFNIRPLIVGPGIKTCLNKLKVDNPRE-----IGSDRIVNAAA 117
QY 120 IENYSFENVLVYDLGTACTIFAVSRQDGIILGGLIINSGLPINFNSLDNAYLIKFKFISTP 179
DB 118 SEEYGT-TPVIVVDFGTATTTCYIDESGVYGGGAIAFGIMISTEALYNRAAKLPVDAES 176
QY 180 NNLLERTTSGVNSGLFYQYKYLIEGYRDIKQYKKNLITGGNADILSLIETIEFI 239
DB 177 SQIIGKSTVSSMQAGIFVGVQCEGIIAEMKKQSNASPVVYVATGGLARMITEKSSAVDI 236
QY 240 FNIHLTVGVRIL 252
DB 237 LDFPFLTKGLELL 249
RESULT 11
AAU91163
ID AAU91163 standard; Protein; 233 AA.
```





AAU91153;  
 05-JUN-2002 (first entry)  
 Rhodobacter capsulatus pantothenate kinase Coax.  
 Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 Rhodobacter capsulatus.  
 WO200216601-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US26531.  
 24-AUG-2000; 2000US-227860P.  
 20-MAR-2001; 2001US-0813453.  
 (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 Yocum RR, Patterson TA;  
 WPI; 2002-269358/31.  
 Identifying potential antibiotic or antimicrobial agent, comprises  
 contacting composition comprising pantothenate kinase (Coax) protein  
 with test compound and identifying inhibitor of the Coax protein -  
 Claim 10; Page 71-72; 128pp; English.  
 The invention describes assays for identifying a (potential) antibiotic  
 comprising contacting an assay composition comprising a pantothenate  
 kinase (Coax) protein with a test compound, and determining the ability  
 of the test compound to inhibit the activity of the Coax protein, an  
 essential enzyme for the production of coenzyme A. Coax protein is a  
 valuable target for identifying bactericidal compounds. Coax modulating  
 agents can be used in an infectious animal model to determine the  
 efficacy, toxicity, or side effects of treatment with such an agent. This  
 is the amino acid sequence of a pantothenate kinase (Coax) protein  
 described in the invention.  
 Sequence 258 AA;  
 Query Match 14.8%; Score 197.5; DB 23; Length 258;  
 Best Local Similarity 23.7%; Pred. No. 8.3e-12;  
 Matches 60; Conservative 55; Mismatches 123; Indels 15; Gaps 6;  
 9 LIIDGNTSIAFALFKDQVNLFIKMTNMLRYDEVYSFFE-----ENFDNKNKVFIS 63  
 3 LCIDCGNTNVSVDGTDFAATWRTATDHRRTADEYFVWLNTLMQLKLGRISEAITIS 62  
 64 SVVPILNETFNKVFISFEKIKPLFGDNLNYDLTFNPKSKDFLLGSDVFANLYAAIENY 123  
 63 STAPRVNLRVLCNRYFDCRPYVVG-KPGCELPVAPRVDPQTGGPDLVNTVAGYDKRH 121  
 124 SFENVLVVDLGTACTIFAVSRODGLIGLINSGLINFNLSLDNAYLIKFFPISTPNLL 183  
 122 G-GDLIVDFGTATTFDVPAGVIGGVIAFGVNLSEALHMAAALPHVDVTPQGV 180  
 184 ERTTSGSVNSGLFYQYKYLIEGVYRDIKOMYKKKFNLIITGGNADLIISLIEIEF-IPN- 241  
 181 GTNTVACIQSGVYMGVIGLIVEIVRQIRMRDRPMKVATG----LASLFDLGLDFDK 236  
 242 --IHLTVGVRL 252  
 237 VEDDLTMHGLRL 249  
 RESULT 14  
 AAU91176  
 ID AAU91176 standard; Protein; 219 AA.

AAU91176;  
 05-JUN-2002 (first entry)  
 Pantothenate kinase (Coax) #14.  
 Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 Dehalococcoides ethenogenes.  
 WO200216601-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US26531.  
 24-AUG-2000; 2000US-227860P.  
 20-MAR-2001; 2001US-0813453.  
 (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 Yocum RR, Patterson TA;  
 WPI; 2002-269358/31.  
 N-PSDB; ABK54197.  
 Identifying potential antibiotic or antimicrobial agent, comprises  
 contacting composition comprising pantothenate kinase (Coax) protein  
 with test compound and identifying inhibitor of the Coax protein -  
 Claim 10; Page 106-107; 128pp; English.  
 The invention describes assays for identifying a (potential) antibiotic  
 comprising contacting an assay composition comprising a pantothenate  
 kinase (Coax) protein with a test compound, and determining the ability  
 of the test compound to inhibit the activity of the Coax protein, an  
 essential enzyme for the production of coenzyme A. Coax protein is a  
 valuable target for identifying bactericidal compounds. Coax modulating  
 agents can be used in an infectious animal model to determine the  
 efficacy, toxicity, or side effects of treatment with such an agent. This  
 is the amino acid sequence of a pantothenate kinase (Coax) protein  
 described in the invention.  
 Sequence 219 AA;  
 Query Match 14.3%; Score 190; DB 23; Length 219;  
 Best Local Similarity 25.0%; Pred. No. 3.8e-11;  
 Matches 58; Conservative 45; Mismatches 87; Indels 42; Gaps 6;  
 1 MNKPLSELIIIDGNTSIAFALFK-----DQVNLFIKMTNMLRY 42  
 1 MSEKLV-----VDIGNTSVNIIGIFEGEKKLLANWHLGSAQMADEYASLLGLQHAGIHP 57  
 43 DEVYSFFEENFDNKNKVFISVVPILNETFNKVFISFEKIKPLFGDNLNYDLTF---N 99  
 58 EE-----LNRVIMCSVVPPLTTTFFEEVFKSYKAAPLVVGAGIKSGVKVRMDN 105  
 100 PKSKDKFLGSDVFANLYAAIENYSPENVLVVDLGTACTIFAVSRODGLIGLINSGLI 159  
 106 PRE-----VGADRIVNAARVLYP-GACIIIVDMGTATTFDTLSEGGVYIGGAIAPGIAT 159  
 160 NFNSLDNAYLIKFFPISTPNLLERTTSGSVNSGLFYQYKYLIEGVYRDIK 211  
 160 SQAIAETSKLPKIEIIRPAKVIKSGNTVSAMQSGIYFGYIGLVEELVRIQ 211  
 RESULT 15  
 AAU91173  
 ID AAU91173 standard; Protein; 260 AA.  
 XX AAU91173;  
 AC



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.1246 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-41

Perfect score: 1229

Sequence: 1 MSFNLIYDQNSACKVAFVR.....LLIHPDLVLLGLNRLILEYV 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCRT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/PCRTUS\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1229	100.0	244	9	US-09-813-453A-41
2	224	18.2	258	9	US-09-813-453A-6
3	221	18.0	257	9	US-09-813-453A-53
4	210.5	17.1	258	9	US-09-813-453A-49
5	208	16.9	255	9	US-09-813-453A-7
6	207.5	16.9	262	9	US-09-813-453A-45
7	207	16.8	258	9	US-09-813-453A-2
8	205.5	16.7	265	9	US-09-813-453A-4
9	201.5	16.4	256	9	US-09-813-453A-55
10	198	16.1	260	9	US-09-813-453A-51
11	197.5	16.1	254	9	US-09-813-453A-47
12	195.5	15.9	248	9	US-09-813-453A-20
13	184	15.0	219	9	US-09-813-453A-57
14	183.5	14.9	241	9	US-09-813-453A-63
15	181.5	14.8	242	9	US-09-813-453A-65
16	181	14.7	249	9	US-09-813-453A-61
17	178	14.5	249	9	US-09-813-453A-70
18	175.5	14.3	233	9	US-09-813-453A-17
19	174	14.2	212	9	US-09-813-453A-59

20	172.5	14.0	273	9	US-09-813-453A-10	Sequence 10, Appl
21	167.5	13.6	250	9	US-09-813-453A-3	Sequence 3, Appl1
22	166	13.5	460	9	US-09-813-453A-39	Sequence 39, Appl
23	163	13.3	267	9	US-09-813-453A-15	Sequence 15, Appl
24	162	13.2	246	9	US-09-813-453A-9	Sequence 9, Appl1
25	161	13.1	592	9	US-09-813-453A-43	Sequence 43, Appl
26	156	12.7	592	9	US-09-813-453A-22	Sequence 22, Appl
27	153	12.4	272	9	US-09-712-363-276	Sequence 276, App
28	153	12.4	272	9	US-09-813-453A-5	Sequence 5, Appl1
29	152	12.4	257	9	US-09-813-453A-13	Sequence 13, Appl
30	135	11.0	262	9	US-09-813-453A-11	Sequence 11, Appl
31	132.5	10.8	229	9	US-09-813-453A-12	Sequence 12, Appl
32	131	10.7	262	9	US-09-813-453A-8	Sequence 8, Appl
33	129.5	10.5	223	9	US-09-895-913A-74	Sequence 74, Appl
34	129.5	10.5	223	9	US-09-813-453A-14	Sequence 14, Appl
35	129.5	10.5	223	9	US-09-813-453A-67	Sequence 67, Appl
36	126.5	10.3	209	9	US-09-813-453A-21	Sequence 21, Appl
37	95.5	7.8	468	10	US-09-815-242-5530	Sequence 5530, Ap
38	95.5	7.8	468	10	US-09-815-242-12355	Sequence 12355, A
39	95.5	7.8	468	10	US-09-815-242-12934	Sequence 12934, A
40	91.5	7.4	783	9	US-09-738-626-5017	Sequence 5017, Ap
41	86	7.0	650	9	US-09-993-241-2	Sequence 2, Appl1
42	86	7.0	650	10	US-09-993-038-2	Sequence 2, Appl1
43	86	7.0	653	9	US-10-206-443-2	Sequence 2, Appl1
44	86	7.0	1140	9	US-09-974-973-19	Sequence 19, Appl
45	86	7.0	1140	9	US-09-738-626-4265	Sequence 4265, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-813-453A-41

; Sequence 41, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE OF INVENTION: ANTIBIOTICS

; FILE REFERENCE: OG2-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 41

; LENGTH: 244

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-09-813-453A-41

Query Match 100.0%; Score 1229; DB 9; Length 244;

Best Local Similarity 100.0%; Pred. No. 3.3e-108;

Mismatches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSFNLIYDQNSACKVAFVRNNSISISFLPKAGKAGLSHLVAPHDFKAIYSSVGLPDE	60
DB	1	MSFNLIYDQNSACKVAFVRNNSISISFLPKAGKAGLSHLVAPHDFKAIYSSVGLPDE	60

QY	61	EAEAIIVRSCAAASLMNGTETVPVLRLOVDRLRRLTGADRLAAVVGAAHSLYPNTELLVLDAGT	120
DB	61	EAEAIIVRSCAAASLMNGTETVPVLRLOVDRLRRLTGADRLAAVVGAAHSLYPNTELLVLDAGT	120

QY	121	AITTVERSAEGIYLGNGNISPGHLRFLKALHFTGRLPLDPSGISPKIAEYSGSSTEAT	180
DB	121	AITTVERSAEGIYLGNGNISPGHLRFLKALHFTGRLPLDPSGISPKIAEYSGSSTEAT	180

QY	181	AGVHGLAGETIDRYIDDLHAKRGAIVLTGCDANYLARIIRSGILIHFDVLLGNRL	240
DB	181	AGVHGLAGETIDRYIDDLHAKRGAIVLTGCDANYLARIIRSGILIHFDVLLGNRL	240

Db 181 AGVIHGLAGEIDRYIDDLHAKGSAVILTGDDANYLARIIRSGILHLPDLVLLGNRL 240  
QY 241 EYNV 244  
Db 241 EYNV 244

## RESULT 2

US-09-813-453A-6  
; Sequence 6, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Rhodobacter capsulatus  
US-09-813-453A-6

Query Match 18.2%; Score 224; DB 9; Length 258;  
Best Local Similarity 29.4%; Pred. No. 2.6e-13;  
Matches 72; Conservative 37; Mismatches 86; Indels 50; Gaps 10;  
QY 18 FVRNNSIESISFLPKAGQALSHLVAP-----HREFKAIYSSVGLPDEEAEIV 66  
Db 40 FVWLNLQKLGQKRISEAISSAPRVFVNLVLCNRYFDCRPY-VVGKP----- 90  
QY 67 RSCAAASLMGTETVPLRLQYDRRTLGADRLAAVVGASHLPNTLLVIDAGTATYER 126  
Db 91 -----GCPLVAPRVD-PGTVGPDRLVNTVAGYDRH-GGDLIVDFGTATFDV 138  
QY 127 VSAEGIYLGNNISPGHLRFKALHFTGRLPLID---PSGISPKIAEYSSSTEEAITAGV 183  
Db 139 VAPDGAYIGGVITAPGVNLSLEALHMAAALPHVDVTKPGV-----IGTNVACIQSGV 192  
QY 184 IHGLAGEIDRYIDDLHAKGSAVILTGDDANYLARIIRSGI-----LIHPLDLVLLGNR 238  
Db 193 YWYIGLVEGIVRQIRMERDRPKVIATGG-----LASLFDLGLDFDKVEDDLTHGLRL 248  
QY 239 ILEYN 243  
Db 249 IFDYN 253

## RESULT 3

US-09-813-453A-53  
; Sequence 53, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Chlorobium tepidum  
US-09-813-453A-53  
Query Match 18.0%; Score 221; DB 9; Length 257;  
Best Local Similarity 28.8%; Pred. No. 4.9e-13;  
Matches 76; Conservative 43; Mismatches 111; Indels 34; Gaps 8;  
QY 5 LIVDQNSACKVAFVRNNSIESISFLPKAGQALSHLVAPRFDKAIYSSVGLPDEEAE 64  
Db 3 LVVDIGNTSTLAIFTGDEEPSVESVP-----SALFADSTSTREVEFGNMARKHGEPOA 55  
QY 65 I-VRSCAAASLMGT-----ETPV-----PLRLQY-DRRTLGADRLAAVVGASH 106  
Db 56 IATCSVVPSTAVGSALESLSFVPLTICCKLRFPELDYATPHTGADRLALCAWSRH 115  
QY 107 LXPNTLLVIDAGTATYERVSAGIYLGNNISPGHLRFKALHFTGRLPLIDPSGISP 166  
Db 116 LFSEKPVIAVDIGTATFDVLTGVNRYRGGLIMPIDMAGALHSRTAQLPQV---RIDR 172  
QY 167 KIAEYSSSTEEAITAGVIHGLAGEIDRYID---DLHAKGERSA---VILTGDDANYLARI 220  
Db 173 PESLLGRSTTECIKSGVFWGVQIGGLVDAIRGDLVDFGSESTVEVITGNGSRILVPE 232  
QY 221 IIRSGILHPLDLVLLGNRLILEYV 244  
Db 233 IGPVSVIDEVLAVLRGSDLLLRMM 256

## RESULT 4

US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 17.1%; Score 210.5; DB 9; Length 258;  
Best Local Similarity 30.9%; Pred. No. 4.9e-12;  
Matches 68; Conservative 37; Mismatches 80; Indels 35; Gaps 10;  
QY 49 KAIYSSVGLPDEEAEAIIVRSCAAASLM-----MGTEVPVPLRLQYDR- 90  
Db 44 KALLNHVQLQSFDIRGIIISVVPPIMFALERMCLKVFIHKLIVGPGIKTGLDKYDNP 103  
QY 91 RTLGLADRLA-AVVGASHLPNTLLVIDAGTATYERVSAGIYLGNNISPGHLRFKAL 149  
Db 104 REVGADRIYNVAGIHL-LY-GSPLIIVDFGTATTCYINEKHKGMAIAPGIMISTEAL 161  
QY 150 HLFTGRLPLID---PSGISPKIAEYSSSTEEAITAGVIHGLAGEIDRYIDDLHAKG- 205  
Db 162 FARAALKPRIETARPDDI-----IGKNTVSAMQAGILYGVGVQVEGIVSRMRAKSKIPP 215  
QY 206 AVILTGDDANYLARIIRSGIL--IHPDLVLLGNRLILEYN 243



;; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
;; TITLE OF INVENTION: ANTIBIOTICS  
;; FILE REFERENCE: OGZ-001  
;; CURRENT APPLICATION NUMBER: US/09/813,453A  
;; CURRENT FILING DATE: 2001-03-20  
;; PRIOR APPLICATION NUMBER: US 60/227,860  
;; PRIOR FILING DATE: 2000-08-24  
;; PRIOR APPLICATION NUMBER: 09/667,569  
;; PRIOR FILING DATE: 2000-09-21  
;; NUMBER OF SEQ ID NOS: 77  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 4  
;; LENGTH: 265  
;; TYPE: PRT  
;; ORGANISM: Streptomyces coelicolor  
US-09-813-453A-4

Query Match 16.7%; Score 205.5; DB 9; Length 265;  
Best Local Similarity 33.5%; Pred. No. 1.5e-11;  
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;  
QY 77 GTETPVLRLQYDRRTLGADRLAAVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLG 136  
Db 96 GVTGVPILTDHPKE-VGADRIINAVAAVELY-GGPAIVVDFTGTAITFDVARSAGEYIG 153  
QY 137 NISPGHLRFPKALHLFTGRPLIDPGISPKIAEYSGSSTEEAITAGVIHGLAGEID---- 192  
Db 154 VIAPGIEISVALGVGAQLKTE---VARPRSVIGKNTVEAMQSGIVYGFAGQYDGVVN 210  
QY 193 RYDDHLHAKGRSAVILTGADNYLARI--RSGL--IHPDLVLGLNRLILEYN 244  
Db 211 RMARELADDDVTVIATGG----LAPMVLGESSVIDEHEPWTLMGLRLVYERN 262

## RESULT 9

US-09-813-453A-55  
Sequence 55, Application US/09813453A  
Patent No. US20020168681A1  
GENERAL INFORMATION:  
APPLICANT: Yocum, R. Rogers  
APPLICANT: Patterson, Thomas A.  
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
TITLE OF INVENTION: ANTIBIOTICS  
FILE REFERENCE: OGZ-001  
CURRENT APPLICATION NUMBER: US/09/813,453A  
CURRENT FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: US 60/227,860  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: 09/667,569  
PRIOR FILING DATE: 2000-09-21  
NUMBER OF SEQ ID NOS: 77  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 55  
LENGTH: 256  
TYPE: PRT  
ORGANISM: Clostridium difficile  
US-09-813-453A-55

Query Match 16.4%; Score 201.5; DB 9; Length 256;  
Best Local Similarity 27.0%; Pred. No. 3.4e-11;  
Matches 75; Conservative 44; Mismatches 94; Indels 65; Gaps 11;  
QY 5 LIVDQGNACKVAFVRNRSI-----ESTSFLPGKAGQALSHLVAPHREDKAIYSS 54  
Db 3 LVFDVGNNTNVLGIYKDKLVNWKTRDKTS---DEYGLISNL-----FD---YDN 51  
QY 55 VGLPDEARAI-----VRSCAASLMMGTETPVLRLQYDR-RTLQADRL 98  
Db 52 VNISDIDDVLISSVWPNVMSHLENFCIKYCKKQPLIVGPGIKTGLNIDPNKQVADRI 111  
QY 99 AAVVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLGNTSPGLHLRFPKALHLFTGRPL 158  
Db 112 VNAVAGIEKYGAPSIIV-DFGTATTCFAISEKGEYLGGTIAPGKITSSEALFOSAKLPR 170

QY 159 IDPSGISPKIAEYG-----SSTEEAITAGVIHGLAGEIDRYI-----DDLHAKEGRS 205  
Db 171 VE-----LAKPGMTICKSTVSAMQSGIIYGVGLVDKIISIMKKELNCDDV----- 216  
QY 206 AVILTGADNYLARIIRSGILIHDPDLVLLGLNRLILEYN 243  
Db 217 KVIATGGGLAKLIASETSIDYVDGFLTLEGLRIIYERN 254

## RESULT 10

US-09-813-453A-51  
Sequence 51, Application US/09813453A  
Patent No. US20020168681A1  
GENERAL INFORMATION:  
APPLICANT: Yocum, R. Rogers  
APPLICANT: Patterson, Thomas A.  
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
TITLE OF INVENTION: ANTIBIOTICS  
FILE REFERENCE: OGZ-001  
CURRENT APPLICATION NUMBER: US/09/813,453A  
CURRENT FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: US 60/227,860  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: 09/667,569  
PRIOR FILING DATE: 2000-09-21  
NUMBER OF SEQ ID NOS: 77  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 51  
LENGTH: 260  
TYPE: PRT  
ORGANISM: Caulobacter crescentus  
US-09-813-453A-51

Query Match 16.1%; Score 198; DB 9; Length 260;  
Best Local Similarity 33.7%; Pred. No. 7.4e-11;  
Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;  
QY 74 LMMGTETPVLRLQYDRRT-LGADRLAAVGAHSLYPNTELLVIDAGTAITYERVSAEGI 132  
Db 86 LVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAVYGP-PLVVIDSGTATTFDVAADGA 144  
QY 133 YLGNSTSPGHLRFPKALHLFTGRPLI---DPSGISPKIAEYSGSSTEEAITAGVIHGLAG 189  
Db 145 FEGGIIAPGINLSMQALHEAAKLPRIAIQRPAG--NRIV--GTDVTSAMQSGVFWGYIS 200  
QY 190 EIDRYIDDLHAKGR-SAVILTGADNYLARIIRSGILIHDPDLVLLGLNRLILEYN 243  
Db 201 LIEGLVARIKAEKGEPMVTIATGCVASLFEGATSDIDHFDSDLTIRGLLEIYRN 255

## RESULT 11

US-09-813-453A-47  
Sequence 47, Application US/09813453A  
Patent No. US20020168681A1  
GENERAL INFORMATION:  
APPLICANT: Yocum, R. Rogers  
APPLICANT: Patterson, Thomas A.  
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
TITLE OF INVENTION: ANTIBIOTICS  
FILE REFERENCE: OGZ-001  
CURRENT APPLICATION NUMBER: US/09/813,453A  
CURRENT FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: US 60/227,860  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: 09/667,569  
PRIOR FILING DATE: 2000-09-21  
NUMBER OF SEQ ID NOS: 77  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 47  
LENGTH: 254  
TYPE: PRT  
ORGANISM: Bacillus halodurans



US-09-813-453A-47

	Query Match	16.18;	Score 197.5;	DB 9;	Length 254;
	West Local Similarity	27.9%;	Pred. No. 8e-11;		
	Matches 60;	Conservative 44;	Mismatches 88;	Indels 23;	Gaps 6;
QY	49 KAIYSVGLPDEAEATIVSCAAASLMWGTE-----TPV-----PLRLQYDR- 90     :   :   :   :   :     :   :   :				
Dd	44 RSLFDHAGLQFQIDIGVISVVPPMFMSLEQMCKKFHYPTMTIIGIKYTNLIKINIDNP 103     :   :   :   :   :     :   :   :				
QY	91 RTGADRLAIVGAHSLYPNTELLVDAGTAITAYRVSABGIYLGGNISPLGHLPFRKAL 150 : :     :   :   :   :   :     :   :   :				
Dd	104 KEYGADRIVNAVAIEY-GYPATVDFGTATTVCCLINEKKQYAGGVIAFGIMISTEALY 162 : :     :   :   :   :   :     :   :   :				
QY	151 LFTGRPLIPDPISPKIAEYGSSTEEAITAGVTHTGLAGEIDRYDIDLHAK-EGRSAVIL 209       :   :   :   :   :     :   :   :				
Dd	163 HRASKLPRIE---IAPFKQVVGNTNDSMQSGIFYGYVSQDGVVKRNKAQAASEPKVIA 219       :   :   :   :   :     :   :   :				
yy	210 TGDGANVLARIIRSGILIHDPVLVLLGNLRILEYNV 244       :   :   :   :   :     :   :   :				
-bb	220 TGGLAKLTGETSIDVIDSFLTKGLQLIYKKNV 254       :   :   :   :   :     :   :   :				

**RESULT 12**

US-09-813-453A-20  
; Sequence 20, Application US/09813453A  
; Patent No. US20020168681A1

```

; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 20
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

```

Query Match	15.9%; Score 195.5; DB 9; Length 248;
Best Local Similarity	30.4%; Pred. No. 1.2e-10;
Matches	80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;
QY	5 LIVDQNSACKVAFVFNNSIESIFLPGKAGQALSHLVAPHRDCKAIYSSVGLPDEAEA 64
Db	1 :       : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
3	LELDGNSLTKRWI-EGAARSAV---GGLAESDDALV-----EQLTSQQALP----- 46
QY	65 IVRSACAAASLMGTEH-----PVP-----LRLOY-DRRTLGADFLAA 100
Db	1 :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
47	-VRACRLVSVRSQETSQSLVARLEQLFPVSALVASSGKQLAGVANGVLDQRLGLDRWLA 105
QY	101 VYGAHSLYPNTELLVDAQTATYERVSABGIVLGGNISPLGRFKALHLFTGRPLID 160
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
106	LVAAHHL-AKKACLVLDLGTAVTSDLVADGVHLGGVYICPGWILMSQLTHRTFRYYDD 164
QY	161 P-----SGISPKIAYGSGSTEAITAGVIHGLAGET-DRYIDDLHAKEGRSAVILTGGD 213
Db	: : : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
165	AEARRALASLP-----GQATAEAVRGCLLMRGVREQYAMACELLGPDCEIFLTGGD 219
QY	214 ANYLARIIRSGILIHDPDLVLGL 236
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
220	AE-LVRDELAGARIMPDLVFVGL 241

RESULT 13

US-09-813-453A-57

```

> Sequence 57, Application US/09813453A
> Patent No. US20020168681A1
> GENERAL INFORMATION:
> APPLICANT: Yocum, R. Rogers
> APPLICANT: Patterson, Thomas A.
> TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
> TITLE OF INVENTION: ANTIBIOTICS
> FILE REFERENCE: OGZ-001
> CURRENT APPLICATION NUMBER: US/09/813,453A
> CURRENT FILING DATE: 2001-03-20
> PRIOR APPLICATION NUMBER: US 60/227,860
> PRIOR FILING DATE: 2000-08-24
> PRIOR APPLICATION NUMBER: 09/667,569
> PRIOR FILING DATE: 2000-09-21
> NUMBER OF SEQ ID NOS: 77
> SOFTWARE: PatentIn Ver. 2.0
> SEQ ID NO 57
> LENGTH: 219
> TYPE: PRT
> ORGANISM: Dehalococcoides ethenogenes
> US-09-813-453A-57

```

Query Match	15.08;	Score 184;	DB 9;	Length 219;
Best Local Similarity	27.08;	Pred. No. 1.2e-09;		
Matches	61; Conservative	42; Mismatches	91; Indels	32; Gaps
				9;
QY	1	MSFNLI-VDOGNSACKVAFVRNNSI-----ESISFLPKAGQAALSHLVAP	44	
		:   :   :		
Db	1	MSEKLVAVDIGNTSVNIGIFEGEKKLANHWLGSVAORMADEYASLLGLQLQHAGIH---	57	
		:   :   :		
QY	45	HREDKAIYSVGVLPL-DEEAAIVRS-CAAAASLMGTETPVLRLQYDR-RTLGDADRLAAY	101	
		:   :   :   :		
Db	58	EELNRVIMCSWPPPLATTFEEVPKSYFKAAPLVVGAGIKSGVKVRMDNPREGVADRIVNA	117	
		:   :   :   :		
QY	102	VGAHSLVPNTLEVLVDAGTATTVERVSAGIYLGGNISPGUHLFRFKALHLFTGRLP--L	158	
		:   :   :   :   :		
Db	1..8	AARVLTPGA-CIIIVDMGTATTTDTLSEGGAYICGATAPGATSAQAIEKTSKPKEI	176	
		:   :   :   :   :		
QY	159	IDPSGISPKIAEYGSSTFEAITAGVIHGLAGEIDRYIDDLHAKEGR	204	
		:   :   :   :   :		
Db	177	IRPAKV-----IGSNIVSAMOSGIYFGIYGLVEELVRRITOTELGO	216	
		:   :   :   :   :		

RESUIT 14

RESOLUTION 14  
US-09-813-453A-63  
; Sequence 63, Application US/09813453A  
; Patent No. US20020168681A1

```

; CURRENT APPLICATION NUMBER: US/09/813,453A
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Thiobacillus ferrooxidans
US-09-813-453A-63

```

Query Match 14.9%; Score 183.5; DB 9; Length 241;  
Best Local Similarity 35.4%; Pred. No. 1.6e-09;  
Matches 51; Conservative 23; Mismatches 65; Indels 5

92 TLGADRLAAVVG AHS LYPNT ELLVIDAGTAITYERSAEGIYLG GNI SPGLHLRFKALHL 151

Db 96 SLGDRRCCLAAADYPGQDSIVIDMGTAITID-LLAGGHFRGRILPGIAMSRLGLHE 154  
QY 152 FTGRLPLIDPSGISPKIAEYGSSTERAITAGVHGLAGEIDRYIDDLHAKEGRSVILTG 211  
Db 155 GTALLPEV---VLNAPAEMLGNDTSNAIQAGVHFLFADALRGAITDPRQYSPQARILITG 211  
QY 212 GDANYLARIIRSGILIHDPDLVLG 235  
Db 212 GDAERWQPGI-AGSLYQPHLLLRG 234

## RESULT 15

US-09-813-453A-65  
; Sequence 65, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 65  
; LENGTH: 242  
; TYPE: PRF  
; ORGANISM: Xylella fastidiosa  
US-09-813-453A-65

Query Match 14.8%; Score 181.5; DB 9; Length 242;  
Best Local Similarity 28.0%; Pred. No. 2.4e-09;  
Matches 72; Conservative 39; Mismatches 91; Indels 55; Gaps 12;  
QY 6 IVDQGNACKVAFVRNNSIESIFLP-----GKAGQALSHLVAPHRFDK 49  
Db 5 LFDLGNRKFCKASLRGIVGVPVLPVLTETMDAFALQELPRGVAY-LASVAAP----- 58  
QY 50 AIVS-----SYGLPDEEAEAVRSCAAASLMMGTETPVLRLQYDR-RTLGAADRLAAVY 102  
Db 59 AITTHVLEVLKIHFEQVQVAATVACAG-----VRIAYAHPERFGVDRELLALL 106  
QY 103 GAHSLYPNTLAVIDAGTAITYERSAEGIYLGNNISPGHLHFRKALHFTGRPLIDPS 162  
Db 107 GS---YGEGNVLVGVGTALTIDLLAANGCHLGGRISSASPTLMRQALHARAQOLPL---S 160  
QY 163 GISPKIAEYGSSTERAITAGVHGLAGEIDRYIDDLHAKEGRSVILTG--GGDANYLARI 220  
Db 161 G--GNYLEAEDETEALVSGCGNGAAVALIERSLYEAHQRLDQSVRLHGGGVASLLPWL 218  
QY 221 IRSGILIH-PDLVLLGL 236  
Db 219 ---GDVHRPTLVLDGL 232

Search completed: June 24, 2003, 22:24:08  
Job time : 12.1246 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: June 24, 2003, 21:35:06 ; Search time 19.9007 Seconds  
(without alignments)  
2526.317 Million cell updates/sec  
Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLVDSGNACKVAEVR.....ILHPLDVLGLNRILEYNV 244  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	223	18.1	295	16 Q8Y2M4	Q8Y2M4 ralstonia s
2	222.5	18.1	258	2 Q9F985	Q9F985 bacillus st
3	216	17.6	259	16 Q8YAC5	Q8YAC5 listeria mo
4	215.5	17.5	255	16 Q8R7M2	Q8R7M2 thermomanaer
5	214	17.4	259	16 Q92F54	Q92F54 listeria in
6	208	16.9	273	16 Q97EB4	Q97EB4 clostridium
7	205.5	16.7	265	16 Q9XBN6	Q9XBN6 streptomyces
8	198	16.1	261	16 Q9A6Z1	Q9A6Z1 caulobacter
9	197.5	16.1	254	16 Q9KGH5	Q9KGH5 bacillus ha
10	195.5	15.9	248	16 Q9HWC1	Q9HWC1 pseudomonas
11	188	15.3	259	16 Q8XHL5	Q8XHL5 clostridium
12	183	14.9	276	16 Q8YQD7	Q8YQD7 anabaena sp
13	181.5	14.8	242	16 Q9PC14	Q9PC14 xyella fas
14	178.5	14.5	256	16 Q8RFE4	Q8RFE4 fusobacteri
15	174	14.2	212	2 Q32514	Q32514 desulfovibr
16	172.5	14.0	273	16 Q83446	Q83446 treponema p

17	162	13.2	246	16 Q9WZY5	Q9WZY5 thermotoga
18	161	13.1	592	16 Q9JXF1	Q9JXF1 neisseria m
19	157.5	12.8	274	16 Q9CD56	Q9CD56 mycobacteri
20	156	12.7	592	16 Q9JW17	Q9JW17 neisseria m
21	153	12.4	272	16 Q96282	Q96282 mycobacteri
22	152	12.4	257	16 P74045	P74045 synecocyst
23	135	11.0	262	16 Q51477	Q51477 borrelia bu
24	134	10.9	223	16 Q92KY6	Q92KY6 helicobacte
25	132.5	10.8	229	16 Q67753	Q67753 aquifex aeo
26	131	10.7	262	16 Q9RX54	Q9RX54 deinococcus
27	129.5	10.5	223	16 Q25533	Q25533 helicobacte
28	126.5	10.3	209	16 Q9PIA9	Q9PIA9 campylobact
29	113	9.2	450	16 Q92Z19	Q92Z19 rhizobium m
30	105.5	8.6	456	2 Q54101	Q54101 saccharopol
31	100.5	8.2	6396	2 Q9KID7	Q9KID7 streptomyce
32	98	8.0	497	16 Q8XW09	Q8XW09 ralstonia s
33	96.5	7.9	385	16 Q9AA68	Q9AA68 caulobacter
34	96	7.8	350	16 Q9R291	Q9R291 deinococcus
35	95.5	7.8	240	16 Q8XBD7	Q8XBD7 escherichia
36	95.5	7.8	300	16 Q92L81	Q92L81 rhizobium m
37	94.5	7.7	534	16 Q8UAV8	Q8UAV8 agrobacteri
38	94.5	7.7	758	2 Q9X6G0	Q9X6G0 lactobacill
39	94	7.6	298	17 Q9HRJ6	Q9HRJ6 halobacteri
40	93	7.6	323	16 Q9CGS8	Q9CGS8 lactococcus
41	93	7.6	559	16 Q9AAK4	Q9AAK4 caulobacter
42	91.5	7.4	342	16 Q8UKM4	Q8UKM4 agrobacteri
43	91.5	7.4	369	16 Q9RD53	Q9RD53 streptomyce
44	91.5	7.4	611	12 Q65855	Q65855 beet yellow
45	91.5	7.4	4077	2 Q93920	Q93920 amycolatops

ALIGNMENTS

RESULT 1

Q8Y2M4	Q8Y2M4	PRELIMINARY;	PRT;	295 AA.
AC	Q8Y2M4;			
DT	01-MAR-2002 (Tremblrel. 20, Created)			
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)			
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)			
DE	Putative transcription regulation accessory factor transcription			
DE	regulator protein.			
GN	RSC0311 OR RS03278.			
OS	Ralstonia solanacearum (Pseudomonas solanacearum).			
OC	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;			
OC	Ralstonia.			
OX	NCBI_TaxID=305;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GM11000;			
RX	MEDLINE=21681879; PubMed=11823852;			
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,			
RA	Arlat M., Billault A., Bottier P., Camus J.C., Cattolico L.,			
RA	Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,			
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,			
RA	Siguer P., Theault P., Whalen M., Wincker P., Levy M.,			
RA	Weissenbach J., Boucher C.A.;			
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum.";			
RL	Nature 415:497-502(2002).			
DR	EMBL; AL646058; CAD13839.1; -.			
DR	InterPro; IPR004619; Baf.			
DR	Pfam; PF03309; Bvg_acc_factor; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;			

Query Match	18.1%;	Score 223;	DB 16;	Length 295;
Best Local Similarity	28.5%;	Pred. No. 2.6e-10;		
Matches	78;	Conservative	38;	Mismatches 106;
				Indels 52;
				Gaps 7;
QY	5	LIVDQGNACKVAFVRNNSIESISFLPG-----KAGQALSHLV-----A	43	
Db	14	LLIDAGNTRIKWATAAD-VAPPVAPGTPWQHAGARPHDQLAELVEDWRDCHAGAWA	72	





```

RESULT 7
Q9X8N6 PRELIMINARY; PRT; 265 AA.
ID AC Q9X8N6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DE 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SC03380.
GN SC03380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmid and a detailed genetic and physical map for
RA the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RA coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL049628; CAB40880.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
DR Hypothetical protein.
RW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 16.7%; Score 205.5; DB 16; Length 265;
Best Local Similarity 33.5%; Pred. No. 6e-09;
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;

QY 77 GTETVPVRLQYDRFTGLGADRLAAVVGASLHPNTLVLLVIDAGTAITVERVSAEGILGG 136
DB 6 GVKTVGPILTDPKRE-VGADRIINAAVELY-GGPAIVDPGTATFDVARSAGEYIGG 153
QY 137 NISPGHLRFKALHLFTGLPLIDPSGISPKIAEYSGSTEEAITAGVIHGLAGEID- 192
DB 154 VIAPGIEISVEALGVKGAQLRKIE---VARPSVIGKNTVEAQMSGIYVGFAGVDGVVN 210
QY 193 RYIDDLHAKGRSAVILGGDANYLARI--RSGIL--IHPDLVLLGLNRLILEYNY 244
DB 211 RWARELADDPDVTVIATGG----LAPVWLGESSVIDEHEPWLTLMLGLRLVYERNV 262

RESULT 8
Q9A6Z1 PRELIMINARY; PRT; 261 AA.
ID AC Q9A6Z1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DE 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.F., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouiri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005867; AAK23910.1;
DR TIGR: CC1935;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs: TIGR00671; baf; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 16.1%; Score 198; DB 16; Length 261;
Best Local Similarity 33.7%; Pred. No. 2.4e-08;
Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;

QY 74 LMGTETVPVRLQYDRFT-IGADRLAAVVGASLHPNTLVLLVIDAGTAITVERVSAEGI 132
DB 87 LVIGENAKLIGDIVRIEKPSEAGADRLVNAIGRAMVYFG-PLVVIDSGTATTFDVAADGA 145
QY 133 YLGGISNPSGLHRLFKALHLFTGLPLI---DPSGISPKIAEYSGSTEEAITAGVIHGLAG 189
DB 146 FEGGIAPGINSMLQALHEAAKLPRIATQRPAG--NRIV--GTDVSAQSGVFWGYIS 201
QY 190 EIDRYIDDLHAKGR-SAVILGGDANYLARIIRSGILIHDPDLVLLGLNRLILEYNY 243
DB 202 LIEGLVARIKAEGRPEMTVIATGGVASLFGATSDIDHFDSDLTIRGLLEIYRRN 256

RESULT 9
Q9KGH5 PRELIMINARY; PRT; 254 AA.
ID AC Q9KGH5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DE 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RA halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AF001507; BAB03805.1;

```

DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRfams; TIGR00671; Baf; 1.  
 KW. Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 16.1%; Score 197.5; DB 16; Length 254;  
 Best Local Similarity 27.9%; Pred. No. 2.5e-08;  
 Matches 60; Conservative 44; Mismatches 88; Indels 23; Gaps 6;

QY 49 KAIYSSVGLPDEAEAIYVRSAAASLMGT-----TPV-----PLRLOYDR- 90  
 DB 44 RSLFDHAGLOFQIDIGIVSSVPPPMFSLQMKCKYFHTPMIIGIKYGLNKKYDNP 103  
 QY 91 RTGADRLAAVGAHSLYPNTLLVIDAGTAIYVRSABGIYLGNSPGLHLRFLKALH 150  
 DB 104 KEVGADRVNVAIAELY-GYPAIVVDGTTATYCLINEKKQYAGGVIAFGIMISTEALY 162  
 Y 151 LFTGRLPLIDPSGISPKIAEYGSSTEEATAGVIHGLAGEIDRYIDDLHAK-EGRSVAIL 209  
 DB 163 HRASKLPRIE---IAKPKVVGVTNTIDSMQSGIFGYVSQVDGVYKRMKAQAESEPKVIA 219  
 QY 210 TGGDANYLARIIRSGILIHDPDLVLGLNRLILEYV 244  
 DB 220 TGSGLAKIGTETIDVIDSFLIKGLQLIYKKNV 254

RESULT 10  
 ID Q9HWC1 PRELIMINARY; PRT; 248 AA.  
 AC Q9HWC1;  
 DT 01-MAR-2001 (TremBLrel. 16, Created)  
 DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE Hypothetical protein PA4279.  
 GN PA4279.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";  
 RL Nature 406:959-964 (2000).  
 DR EMBL; AE004843; BAG07667.1; -;  
 DR InterPro; IPR004619; Baf.  
 DR InterPro; IPR001230; Presyl\_site.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;

Query Match 15.9%; Score 195.5; DB 16; Length 248;  
 Best Local Similarity 30.4%; Pred. No. 3.6e-08;  
 Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;

QY 5 LIYDQNSACKVAFVRNNSSTESIFLPGKAGQALSHLVAPRFDKAIYSSVGLPDEAEA 64  
 DB 3 LEIDCGNSLIKRVII-EGAARVA---GGLAESDDALV-----EQITSQALP----- 46  
 QY 65 IVRSCAAASLMGTET-----PVP-----LRLQY-DRRTIGADRLAA 100  
 DB 47 -VRACRLVSRSEQTSQLVARLEQFPVVALVSSGKQAGVYRNGYGLRQLDRWLA 105

QY 101 VVGAHSLYPNTLLVIDAGTAIYVRSABGIYLGNSPGLHLRFLKALHFTGRPLD 160  
 DB 106 LVAAHHL-AKKACLVLDIGTAVTSDLVAAADGVHGGYICPGMTLMRSQRLTHTRRYDD 164  
 QY 161 P-----SGISPKIAEYGSSTEEATAGVIHGLAGEI-DRYIDDLHAKEGRSVAILTGGD 213  
 DB 165 AEARRALASLP-----QATAEAVERGCLLMRLRGVREOYANACELLGPDCEIFLTGGD 219  
 QY 214 ANYLARIIRSGILIHDPDLVLGL 236  
 DB 220 AE-LVRDELACARIMPDLVFVGL 241

RESULT 11  
 ID Q8XHL5 PRELIMINARY; PRT; 259 AA.  
 AC Q8XHL5;  
 DT 01-MAR-2002 (TremBLrel. 20, Created)  
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE Hypothetical protein CPE2468.  
 GN CPE2468.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).  
 DR EMBL; AP003194; BAB82174.1; -;  
 DR InterPro; IPR004619; Baf.  
 DR InterPro; IPR000515; BPD\_transp.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRfams; TIGR00671; Baf; 1.  
 DR PROSITE; PS00402; BPD\_TRANSPIINN\_MEMBER; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A1A5E10 CRC64;

Query Match 15.3%; Score 188; DB 16; Length 259;  
 Best Local Similarity 29.8%; Pred. No. 1.6e-07;  
 Matches 65; Conservative 42; Mismatches 79; Indels 32; Gaps 12;

QY 44 PHREDKAIYSSVGLPD--EEAEIVRSC-AAASLMGTETPVPLRLOYDR-RTLGAADRLA 99  
 DB 54 PEDVEGIITSSV-VPNIMHSLNVMVKCFCKEPIVVGPIKGTGINKYDNPKEVGADRV 112  
 QY 100 AVCAHSLYPNTLLVIDAGTAIYVRSABGIYLGNSPGLHLRFLKALHFTGRLPLT 159  
 DB 113 NAAFAFEKH-KKPMIIDFGTATTTCATTEKGYLGNGICPGIQISADALFAAKLPRI 171  
 QY 160 D---PSGISPKIAEYGSSTEEATAGVIHGLAGEIDRYID-----DLHAKEGRSVAIL 209  
 DB 172 ELERPKSVICK-----NTVTSMQAGIYGVIGKVEYIVKRMKKEMMDLGEKE--PFVLA 223  
 QY 210 TGGDANYLARIIRSGILIHDP--LVLLGLNRLILEYV 243  
 DB 224 TGG----LAKLVSETDVIDEVDRKRLTLEGLKILYEKN 257

RESULT 12  
 ID Q8YQD7 PRELIMINARY; PRT; 276 AA.  
 AC Q8YQD7;  
 DT 01-MAR-2002 (TremBLrel. 20, Created)  
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
 DE Hypothetical protein Alr3896.

```

GN ALR3896.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2159285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
  Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
  Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
  Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
  Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
  cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003594; BAB75595.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 30272 MW; 2AD04CA693E56F25 CRC64;

Query Match 14.9%; Score 183; DB 16; Length 276;
Best Local Similarity 31.3%; Pred. No. 4.4e-07;
Matches 52; Conservative 33; Mismatches 71; Indels 10; Gaps 4;

QY 82 VPLRLQYDRTLGADRLAUVGAHSLYPNTLIVDAGTAITYERSAGIYLGNNISPG 141
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110 IFLNNIYP--TLGIDRALALWAGMSW-GFPVLIVDAGTALTAAADGKGNLWGGAILPG 166
QY 142 LHLRFKALHLFTGRPLIDPSGISPKIAEYSGSTEEAITAGVTHGLAGEIDRYIDLHAK 201
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 VGLQFASLGGTQGLPQVMEAIKLPFPALNTAEIOSGVIIYTLIAGMRDFTTEWLSL 226
QY 202 EGRSAVITGGD---ANYLARI---IRSGILHPLDLVLLGLNRL 240
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 PPDGKVAIKGGDRILLNLYQALYPLDLAARLIVEPNLIWFGQTIV 272

RESULT 13
Q9PC14 PRELIMINARY; PRT; 242 AA.
AC Q9PC14;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DE Hypothetical protein Xf1795.
GN Xf1795.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
EX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
  Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
  Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
  Bueno M.R.P., Canargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
  Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
  Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Doory H.,
  Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
  Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
  Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
  Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
  Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
  Lenos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
  Machado M.A., Martins E.A.B.N., Madeira H.M.F., Marino C.L.,
  Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
  Moon D.H., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
  Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
  de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

```

```

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
  Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
  de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
  da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
  da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
  de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
  Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
  Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004001; AAF84603.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 242 AA; 25658 MW; EDC5DE7C049C1E9 CRC64;

Query Match 14.8%; Score 181.5; DB 16; Length 242;
Best Local Similarity 28.0%; Pred. No. 4.8e-07;
Matches 72; Conservative 39; Mismatches 91; Indels 55; Gaps 12;

QY 6 IVQDGSACKVAFVRNNSIESISFLP-----GKAGQALSHLVAPHRFDK 49
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 LFDLGNRSFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAY-LASVAAP----- 58
QY 50 AIYS-----SVGLPDEEAIAIVRSACAAASLMGTEPTVPLRLOYDR-RTLGADRLAUV 102
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
59 AIITHVLEVLKIHFEQVQVAATVAACAG-----VRIAYAHPEFGVDRLALL 106
QY 103 GAHSLYPNTLIVDAGTAITYERSAGIYLGNNISPGILHLRFKALHLFTGRPLIDPS 162
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 GS---YEGNVLVVGVTALTIDLLANGCHLGRISASPTLMQALHARAQLPL---S 160
QY 163 GISPKIAEYSGSTEEAITAGVTHGLAGEIDRYIDLHAKGRSAVILT--GGDANYLARI 220
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 G--GNYLEPAEDTFEDALVSCNGAAVALIERSLEYAHRDQSVRLLLHGGVSLPWL 218
QY 221 IRSGILH-PLDLVLLGL 236
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219 --GDVVHRTPLVDGL 232

RESULT 14
Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 25586;
EX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
  Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
  Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
  Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
  Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
  nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010586; AAL94957.1; -.
KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 14.5%; Score 178.5; DB 16; Length 256;
Best Local Similarity 31.8%; Pred. No. 9.2e-07;
Matches 49; Conservative 27; Mismatches 67; Indels 11; Gaps 4;

```



**RESULT 15**

Search completed: June 24, 2003, 21:59:41  
Job time : 20.9007 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.128 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIYDQGSACKVAFVR.....LLIHPDLVLGLNRIEYV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	7.5	473	4	US-09-134-001C-3405
2	87	7.1	815	4	US-09-134-001C-3614
3	86	7.0	653	2	US-08-713-928B-9
4	86	7.0	653	4	US-09-191-171-5
5	86	7.0	653	4	US-09-385-707-5
6	86	7.0	653	4	US-09-439-923-2
7	86	7.0	1140	4	US-09-220-081-2
8	86	7.0	1140	4	US-09-677-575-2
9	83	6.8	467	4	US-09-129-112-19
10	79	6.4	741	4	US-09-001-984C-106
11	79	6.4	6095	4	US-09-144-085-2
12	78.5	6.4	271	4	US-09-318-794A-2
13	78.5	6.4	271	4	US-09-318-793A-2
14	77.5	6.3	3739	3	US-09-320-878-2
15	77	6.3	554	4	US-08-462-467B-22
16	77	6.3	983	2	US-08-164-292B-26
17	77	6.3	983	3	US-08-845-623-26
18	77	6.3	983	3	US-08-815-927-26
19	77	6.3	983	4	US-09-103-330-26
20	77	6.3	983	4	US-09-435-242-26
21	77	6.3	1399	4	US-08-462-467B-14
22	76.5	6.2	688	1	US-07-688-352C-28
23	76.5	6.2	688	2	US-08-474-379C-28
24	76.5	6.2	688	3	US-09-146-249A-28
25	76.5	6.2	688	3	US-08-206-188B-28
26	76	6.2	461	4	US-09-355-115-7
27	76	6.2	1039	4	US-09-409-648-7

28	76	6.2	1039	4	US-09-409-648-8	Sequence 8, Appli
29	76	6.2	1039	6	5196511-2	Patent No. 5196511
30	75.5	6.1	284	4	US-09-134-001C-2927	Sequence 2927, Ap
31	75	6.1	255	1	US-08-225-757B-2	Sequence 2, Appli
32	75	6.1	255	2	US-08-722-050-2	Sequence 2, Appli
33	75	6.1	415	4	US-09-182-450-1	Sequence 1, Appli
34	74	6.0	4536	4	US-09-180-422B-27	Sequence 27, Appli
35	73.5	6.0	529	4	US-08-887-534A-74	Sequence 74, Appli
36	73.5	6.0	3739	4	US-09-105-537-33	Sequence 33, Appli
37	73.5	6.0	11877	4	US-09-105-537-32	Sequence 6, Appli
38	73	5.9	993	1	US-08-444-792-2	Sequence 2, Appli
39	73	5.9	993	1	US-08-445-042-2	Sequence 53, Appli
40	72.5	5.9	480	4	US-09-004-838-53	Sequence 53, Appli
41	72.5	5.9	531	4	US-09-134-001C-4920	Sequence 4920, Ap
42	72.5	5.9	5087	4	US-09-144-085-1	Sequence 1, Appli
43	71.5	5.8	441	1	US-08-403-866-10	Sequence 10, Appli
44	71.5	5.8	2152	4	US-09-036-987A-3	Sequence 3, Appli
45	71.5	5.8	2152	4	US-09-370-700-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-134-001C-3405

; Sequence 3405, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3405

; LENGTH: 473

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3405

Query Match 7.5%; Score 92; DB 4; Length 473;

Best Local Similarity 24.9%; Pred. No. 0.098;

Matches 46; Conservative 22; Mismatches 51; Indels 66; Gaps 9;

QY 47 FDKATYSSV--GLPDEEAEIVRSCAAASLMGTETPPVLRLOY----DRRTLGADRLAA 100

DB 220 FEKQMTQPVKGMKEGIEIVTEAMAKSA-----EETENGKVKYTEAKGEQTIEADYVLV 275

QY 101 VVGHSLYPNTE-----LIVIDAGTAITYRVRSAGIYLGGNISPGIHLR 145

DB 275 TVGRR--PNTDELGLEELGKFDAGRLLEVD-----KQSRISIENIFAIGDIVPGLPLA 327

QY 145 FKALHLFTGRPLIDPSGISPKIARYGSTERAIYAGVIHGLAGEIDRIDDLHAKGRS 205

DB 328 HKA-----SYEGKVAEAEIDGQAAEVD-YI-----GMP 354

QY 206 AVILT 210

DB 355 AVCFT 359

RESULT 2

US-09-134-001C-3614

; Sequence 3614, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3614  
LENGTH: 815  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3614

Query Match 7.1%; Score 87; DB 4; Length 815;  
Best Local Similarity 20.9%; Pred. No. 0.82;  
Matches 69; Conservative 32; Mismatches 101; Indels 128; Gaps 12;

QY 4 NLIVDGNACKVAFVRRNNSIESIFLPGKAGQALSHLVAPHFRDKAIYSSVGLPDEAE 63  
Db 117 NIWVTUGENAINDLTRPEILESVPSPKPK-----QKQIADQINANQSEK 163  
QY 64 AIVR-----SCAAASLMGTETPV-----PLRLQYDRRLG---ADRLAAYVG--- 103  
Db 164 IMIRLHDLGFGPKLSMAIYQFYMGDTLNVLDKNPYQLVYDIKIGFKNKADQLARNVGIEP 223  
QY 104 -----AHSLYPNT-----ELLVIDAGTAITYE----- 125  
Db 224 HSPERLKAALLTLEECIKQGHYLPRTIVETTONLLNEDIEKPIETEQLLEIDVLS 283  
QY 126 -----RVSAGIYLGNIISPLGLHLRFKALHLFTGRPLIDPSPGISPKIAE----- 170  
Db 284 EEKLLISEADQVSIPSLYSYSELKSVQNLRYKT---NTSKLEIQSDLIQHIGDIESQN 340  
QY 171 ---YGSSTEATAGV-----IHGLAGEIDRYIDD--- 197  
Db 341 EYNSASQKEALETAINSKIMLLTGGPGTGKTVIKGIVELXAEIHGLSLDYDDYEDDY 400  
QY 198 ---LHAKEGRSAVIL---TGGDANYLARI 221  
Db 401 PVVLAAPTGRASKRLHESITGLEAMTHRLI 430

RESULT 3  
US-08-713-928B-9  
Sequence 9, Application US/08713928B  
Patent No. 5929304  
GENERAL INFORMATION:  
APPLICANT: RADIN, DAVID N.  
APPLICANT: CRAMER, CAROLE L.  
APPLICANT: OISHI, KAREN K.  
APPLICANT: WEISSENBORN, DEBORAH L.  
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN  
TITLE OF INVENTION: PLANT-BASED EXPRESSION SYSTEMS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/713,928B  
FILING DATE: 13-SEP-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3614  
LENGTH: 815  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3614

Query Match 7.1%; Score 86; DB 2; Length 653;  
Best Local Similarity 20.7%; Pred. No. 0.75;  
Matches 53; Conservative 39; Mismatches 100; Indels 64; Gaps 10;

QY 25 ESTSFLPGKAGQALSHLVAPHFRDKAIYSSVGLPDEE-----AEAIVRSCA 70  
Db 243 DGTNFTGEAGVRLDY-ISLHR--KGARSSISILEQEKVVAQEIQLRFPKFDPTIYNDE 299  
QY 71 AASLMGTETPVPLRLQYDRRLTGADRLAAYVGAHSLYPNTTELLVIDAGTAITYERSAE 130  
Db 300 ADPL-VGWSLPQPMW-----ADVTYAAVVYKVIQAONLALLANTTSAPFYALLSND 349  
QY 131 GIVL-----GGNISPLGLHLRFKALHLFTGRPLIDPSPGISPKIAEY 172  
Db 350 NAFSLYHPHFAORTLTARQVNVNTRPPHVLRLKPKVLTAMGLALLDEEQLNAEVSQAG 409  
QY 173 SSTEAITAGVI---HGLAGEIDRY-----IDDLHAKEGRSAVIL-----TGDA 214  
Db 410 TVLDSNHTVGLASAHRRPQGPADAWRAAVLIYASDDTRAHPNRSVAVTLRLRGVPPGPG 469  
QY 215 NYLARIIRSGILHDP 230  
Db 470 VVTRYLDNG-LCSPD 484

RESULT 4  
US-09-191-171-5  
Sequence 5, Application US/09191171  
Patent No. 6149909  
Patent No. 6149909 6143294  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

**STREET: 400 Garden City Plaz**

; SEQ ID NO 2









[illegible]

Search completed: June 24, 2003, 21:49:41  
Job time : 8.128 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.24384 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-41

Perfect score: 1229

Sequence: 1 MSFNLIIVDGNACKVAFVR.....ILHPDLVLLGLNRILEYNV 244

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	175.5	14.3	233	1 YACB_BACSU	P37564 bacillus su
2	163	13.3	267	1 BAF_BORPE	Q45338 bordetella
3	95.5	7.8	232	1 YPFH_ECOLI	P76561 escherichia
4	95.5	7.8	468	1 ALDH_STAAM	Q59822 staphylococ
5	95	7.7	332	1 ALF_TREPA	O83668 treponema p
6	90.5	7.4	758	1 CHLD_TOBAC	O24133 nicotiana t
7	88	7.2	417	1 U183_ARATH	O49d33 arabidopsis
8	88	7.2	555	1 FPRE_MYCLE	Q33084 mycobacteri
9	87.5	7.1	449	1 GSHR_BURCE	P48639 burkholderi
10	86	7.0	653	1 IDUA_HUMAN	P35475 homo sapien
11	95	6.9	389	1 AMHX_BACSU	P54983 bacillus su
12	84.5	6.9	941	1 GCSP_MYCTU	Q50601 mycobacteri
13	84	6.8	357	1 PYRD_MYCTU	O06236 mycobacteri
14	84	6.8	612	1 LULB_LYCPN	O04974 lycopersico
15	83.5	6.8	242	1 FABG_ACTAC	P70720 actinobacil
16	83.5	6.8	710	1 NASC_BACSU	P42434 bacillus su
17	83.5	6.8	1149	1 HDA6_MOUSE	Q922v5 mus musculu
18	83	6.8	360	1 BUK_ENTFA	Q9IPs7 enterococcu
19	83	6.8	422	1 PUR2_MYCLE	Q50144 mycobacteri
20	82.5	6.7	428	1 ORDL_RHIME	Q923s3 rhizobium m
21	82	6.7	288	1 HEMK_MYCLE	P45832 mycobacteri
22	82	6.7	405	1 CL30_MYCTU	Q11062 mycobacteri
23	82	6.7	816	1 SUS2_ORYSA	P31924 oryza sativ
24	81.5	6.6	313	1 MSRW_SHEVI	Q9FIN8 shewanella
25	81.5	6.6	612	1 LLVD_HAEIN	P44851 haemophilus
26	81	6.6	274	1 CCS_HUMAN	Q14618 homo sapien
27	81	6.6	274	1 CCS_RAT	Q9J7K2 rattus norv
28	81	6.6	326	1 LEU3_ARCFU	O29627 archaeoglob
29	81	6.6	339	1 TCMQ_STRGA	P39896 streptomyce
30	81	6.6	430	1 TRPB_HALN1	Q9HSC0 halobacteri
31	81	6.6	561	1 MERA_PSEAE	P00392 pseudomonas
32	80.5	6.6	255	1 YJIL_ECOLI	P39383 escherichia
33	80.5	6.6	610	1 DNAK_METTE	Q9UXR0 methanosarc

34 80.5 6.6 983 1 Y644\_MYCLE  
35 80 6.5 331 1 SUCA\_SCHPO  
36 80 6.5 429 1 PUR2\_VIBCH  
37 80 6.5 600 1 RUBB\_ARATH  
38 80 6.5 788 1 TRSL\_HCMVA  
39 79.5 6.5 295 1 YUST\_BACSU  
40 79 6.4 492 1 FYSA\_HELPY  
41 79 6.4 494 1 VYTB\_PLAFA  
42 79 6.4 578 1 PTFB\_RHOCA  
43 79 6.4 598 1 LEPA\_HAEIN  
44 79 6.4 598 1 LEPA\_PASMU  
45 79 6.4 741 1 MASZ\_MYCTU

## ALIGNMENTS

### RESULT 1

YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT: 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=36051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Kobayashi Y., Koetter P., Koningssting G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche H., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,  
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





```

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Magnesium-chelatase subunit chLD, chloroplast precursor (Mg-
DE protoporphyrin IX chelatase) (Mg-chelatase subunit D).
GN CHLD.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=98079235; PubMed=9418040;
RA Papenbrock J., Graefe S., Kruse E., Haenel F., Grimm B.;
RT "Mg-chelatase of tobacco: identification of a Chl D cDNA sequence
RT encoding a third subunit, analysis of the interaction of the three
RT subunits with the yeast two-hybrid system, and reconstitution of the
RT enzyme activity by co-expression of recombinant CHL D, CHL H and CHL
RT I.";
RL Plant J. 12:981-990(1997).
CC CC -I- FUNCTION: Involved in chlorophyll biosynthesis; introduces a
CC magnesium ion into protoporphyrin IX to yield Mg-protoporphyrin
CC IX.
CC CC -I- PATHWAY: Chlorophyll biosynthesis.
CC CC -I- SUBCELLULAR LOCATION: Chloroplast stroma.
CC CC -I- SIMILARITY: BELONGS TO THE MG-CHELATASE SUBUNITS D/I FAMILY.
CC CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL; Y10022; CAA71128.1; -.
CC InterPro; IPR000523; Mg.chelatase_chII.
CC InterPro; IPR002035; VWFA_A.
CC Pfam; PF01078; Mg.chelatase; 1.
CC SMART; SM00327; VWFA; 1.
CC PROSITE; P50234; VWFA; 1.
CC Photosynthesis; Chlorophyll biosynthesis; Chloroplast;
CC Transit peptide.
CC TRANSIT 1 62 CHLOROPLAST (POTENTIAL).
CC CHAIN 63 758 MAGNESIUM-CHELATASE SUBUNIT CHLD.
CC DOMAIN 556 752 VWFA.
CC FT DOMAIN 399 445 GLU/PRO-RICH.
CC FT DOMAIN 409 416 POLY-PRO.
CC FT DOMAIN 427 434 POLY-GLU.
CC FT SEQUENCE 758 AA; 83012 MW; ED31E85D5FAFE8 CRC64;
CC
CC Query Match 7.4%; Score 90.5; DB 1; Length 758;
CC Best Local Similarity 22.8%; Pred. No. 7, 1;
CC Matches 49; Conservative 32; Mismatches 59; Indels 75; Gaps 11;
CC
QY 83 PURLQYDRFTLGADRLAAVVGASHLPNTPELLVID---AGTAITVERVSAEGLYGNIS 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 PEKISGRGYF---PLAAVIGDQAIKATALLGDAIDREIGGATCGKRGTAKTLL---MA 121
QY 140 PGLHLRFKALHLFTGRLPLIDPS-----GTSPKTAEYGS----- 173
Db 122 RGLHALPLPIEVVVGSMANADPCDEWEDGLADR-AEYSGDNKTKQIVKSPFVQPLG 180
QY 174 -----STEEAITAGVI---HGLAGEIDR---YIDDLHAKEGRSAVILTTGGDANYL 217
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 VTEDRLIGSVDYEEVSKSGTTVFQGLLAAHAGRVLYVDEIN-----LLDEGISNLL 232
QY 218 ARIIRSGILI-----HPDLVLGLNRLILEYN 243
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 LNLVTEGVNIVREGISFRHPCKPLP-----IATYN 263

```

```

RESULT 7
U183_ARATH
ID U183_ARATH STANDARD; PRT; 417 AA.
AC Q9SD33; Q94C31;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Putative UPF0183 protein.
DE Putative UPF0183 protein.
GN AT3G51130 OR F24M12.170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
A Salancoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
A Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
A Delaney M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
A De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
A Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
A Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
A Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
A Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
A Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
A Conrad A., Hornischer K., Kauer G., Ioehner T.-H., Nordsiek G.,
A Reichelt J., Schaffe M., Schoen O., Barques M., Terol J., Climent J.,
A Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
A Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
A de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
A Monfort A., Argirou A., Flores J.M., Liguori R., Vitale D.,
A Mannhaupt G., Haase D., Schoof H., Radd S., Zaccaria P., Mewes H.-W.,
A Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
A Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
A Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
A Pai G., Millischer J., Sellers P., Gill J.E., Feldblyum T.V.,
A Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
A Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
A Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
A Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
A Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
A Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana."
RL Nature 408:820-822(2000).
(2)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
SSP consortium (Salk/Stanford/PGSC).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE UPF0183 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL132980; CAB62635.1;
CC DR EMBL; AY037194; AKX59779.1;
CC DR InterPro; IPR0053373; UPF0183.
CC DR Pfam; PF03676; UPF0183; 1.
CC KW Hypothetical protein.
CC CONFLICT 398 417 HFCNNPRFSNFDPCVILADC -> KNCHIATITLFS (IN
FT REF. 2).
SQ SEQUENCE 417 AA; 46629 MW; 7D3460D65E78B8EE CRC64;

```

```

Query Match 7.2%; Score 88; DB 1; Length 417;
Best Local Similarity 22.5%; Pred. No. 5.6;
Matches 34; Conservative 26; Mismatches 51; Indels 40; Gaps 7;
QY 84 LRLOYDRTLGA-DRLAAVVGASLYPNTELVIDAGTAITVERVSAEGIYLGNISPL 142
DB 109 LOMRYATSMIGPSTLATFVAIVLFGPTFPFIYD-----KERGIY--SLFYPGL 156
QY 143 HLRFKALHFT-----GRPLIDPSGISP---KIAEYSSSTEEAITAGVIHGLAGEID 192
DB 157 SFEFIPNQYTDCCHDGEAALPLEPFDGTPVTCRVSIYDSSDKKVGKGL-----MD 210
QY 193 R-----YIDDLHAKEGRSVILFTGG 212
DB 211 RASVPLPPPGSLYMEEVHVKPGKELYFTVGG 241
RESULT 8
FPRB_MYCLE
ID FPRB_MYCLE STANDARD; PRT; 555 AA.
AC Q33064;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Probable ferredoxin/ferredoxin--NADP reductase (EC 1.18.1.2) (FNR).
GN FPRB OR ML2134 OR MLCB57.39.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmsier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K.M., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) -> oxidized
ferredoxin + NADPH.
CC -!- COFACTOR: FAD; PROBABLY BINDS ONE OR TWO 4FE-4S CLUSTERS.
CC -!- SIMILARITY: IN THE N-TERMINAL, BELONGS TO THE BACTERIAL TYPE
FERREDOXIN FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL, TO OTHER FERREDOXIN NADP
REDUCTASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z99494; CAB16679.1;
CC DR EMBL; AL583924; CAC31089.1;
CC DR HSP; P55907; 1XER.
CC DR Leproma; ML2134;
CC DR InterPro; IPR001450; 4Fe4S_ferredoxin.
CC DR Pfam; PF00037; fer4; 2.
CC DR PRINTS; PR00353; 4FE4SFRDOXIN.
CC DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
CC KW Oxidoreductase; Flavoprotein; NADP; FAD; Electron transport;
KW Iron-sulfur; 4Fe-4S; Complete proteome.
FT DOMAIN 1 83 FERREDOXIN.
FT DOMAIN 115 555 FERREDOXIN--NADP REDUCTASE.

```

```

FT METAL          9      9      IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL        15     15     IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL        19     19     IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        46     46     IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        49     49     IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        52     52     IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL        56     56     IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE    555 AA; 59712 MW; 43C7292E6A2DFED9 CRC64;

Query Match          7.2%; Score 88; DB 1; Length 555;
Best Local Similarity 25.4%; Pred. No. 7.8;
Matches 51; Conservative 24; Mismatches 58; Indels 68; Gaps 12;

QY 29 FLPGAGQALSH--LVAPHRFKAIYSSVGLPDEAEALVRSAAASLMGTETPTVPLRL 86
DB 182 FLNVEIGRHLSDLELAHH--AVLYAVGAPDDR-----RU 215

QY 87 QYDRR-----TLGADRLAAVGAHSLYPNTLVIDAGTAITYERSAEGIYLGNI SPGL 142
DB 216 NIDGMCIGPTGTATELVAVINAH---PDFAYLPVD-----LSHERVVVIG---NGNVA--- 262

QY 143 HLRFKALHFTGRLLIDPSGIS-PKIAEY-----GSSTEEAITA---GVHGLAGEID 192
DB 263 -----LDVARLLTADPDNLARTDISEFALHVLGSSAVREVVVAARGPAHSF----- 310

QY 193 RYIDDLHAKEGRSAVILGGD 213
DB 311 -TLPELIGLKATSEVVLDA GD 330

RESULT 9
GSHR_BURCE
ID GSHR_BURCE STANDARD; PRT; 449 AA.
AC P48639;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutathione reductase (EC 1.6.4.2) (GR) (Grase).
GN GOR.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AC1100;
RX MEDLINE=95266809; PubMed=7538273;
RA Daubaras D.L., Hersberger C.D., Kitano K., Chakrabarty A.M.;
RT "Sequence analysis of a gene cluster involved in metabolism of 2,4,5-
trichlorophenoxyacetic acid by Burkholderia cepacia AC1100.";
Appl. Environ. Microbiol. 61:1279-1289(1995).
CC -!- CATALYTIC ACTIVITY: NADPH + oxidized glutathione = NADP(+) + 2
glutathione.
CC -!- COFACTOR: FAD.
CC -!- PATHWAY: 2,4,5-trichlorophenoxyacetic acid degradation.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
OXIDOREDUCTASES CLASS-I.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19883; AAC43334.1; -.
DR HSSP; P28593; 1A0G.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR001100; Pyr_redox.

```

```

DR InterPro; IPR004099; pyr_redox_dim.
DR Pfam; PF00070; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dim; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRODOM; PD000139; FAD_pyr_redox; 1.
DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
KW Redox-active center; Oxidoreductase; Flavoprotein; FAD; NADP.
FT NP_BIND 7 37 FAD (ADP PART) (PROBABLE).
FT DISULFID 43 48 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 292 302 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 435 435 BY SIMILARITY.
SQ SEQUENCE 449 AA; 47541 MW; 402FCC6E7A8D6720 CRC64;

Query Match          7.1%; Score 87.5; DB 1; Length 449;
Best Local Similarity 22.7%; Pred. No. 6.7;
Matches 60; Conservative 35; Mismatches 82; Indels 87; Gaps 14;

QY 6 IVDQNSACKVAFVRNRSIESIFLPCKAGQALSHLVAPHR-----FKAIYSSVGLP 58
DB 172 IIGGGYIACEFAGFNG-----LGR---HVQLHRGSQLRGFDDELREHLG-- 215

QY 59 DE-----EAEIVRSAAASLMGTETPTVPLRLQYDRRTLGADRLAAVGAHS 106
DB 216 DELKSGIDRLRGVDVVAVVERQRCALSVQLTTC-----DAMEVDVMAATG 261

QY 107 LYPNT---ELLVIDAGT---AI---TYERVSAGIYLGNI SPGLHLRFKALH----- 150
DB 262 RLPNTWGLGLETVDGLDQNGAIKVDSEYTSSTSPGIYAVGVDVNTLTPVAIHEGHAF 321

QY 151 --LFTGRLLIDPSGI-----SPKIAEYGSSTEEAITAGVHGLAGEIDRYD-DLHAK 201
DB 322 DTVFGSKALPTEHENVPFAVFSQPAASVGLSEAQA-----RDYRSNVEIYGS 369

QY 202 EGR-SAVILTGGDANYLARIISG 224
DB 370 AFRPMRAALSGRDEKALVKLVNG 393

RESULT 10
IDUA_HUMAN
ID IDUA_HUMAN STANDARD; PRT; 653 AA.
AC P35475;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-iduronidase precursor (EC 3.2.1.76).
GN IDUA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=92052158; PubMed=1946389;
RA Scott H.S., Anson D.S., Orsborn A.M., Nelson P.V., Clements P.R.,
RA Morris C.P., Hopwood J.J.;
RT "Human alpha-L-iduronidase: cDNA isolation and expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:9695-9699(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92372031; PubMed=1505961;
RX Scott H.S., Guo X.H., Hopwood J.J., Morris C.P.;
RT "Structure and sequence of the human alpha-L-iduronidase gene.";
RL Genomics 13:1311-1313(1992).
RN [3]
RP REVIEW ON VARIANTS.
RX MEDLINE=96287378; PubMed=8680403;
RX Scott H.S., Bunge S., Gal A., Clarke L.A., Morris C.P., Hopwood J.J.;
RT "Molecular genetics of mucopolysaccharidosis type I: diagnostic,
clinical, and biological implications.";
Hum. Mutat. 6:288-302(1995).

```



[4]  
 RN VARIANTS MPS-I THR-75.  
 RP MEDLINE-94290500; PubMed-8019563;  
 RX Clarke L.A., Nelson P.V., Warrington C.L., Morris C.P., Hopwood J.J.,  
 RA Scott H.S.;  
 RT "Mutation analysis of 19 North American mucopolysaccharidosis type I  
 RT patients: identification of two additional frequent mutations.";  
 RL Hum. Mutat. 3:275-282(1994).  
 RN [5]  
 RP VARIANTS MPS-I PRO-82.  
 RX MEDLINE-94004880; PubMed-8401515;  
 RA Clark L.A., Scott H.S.;  
 RT "Two novel mutations causing mucopolysaccharidosis type I detected by  
 RT single strand conformational analysis of the alpha-L-iduronidase  
 RT gene.";  
 RL Hum. Mol. Genet. 2:1311-1312(1993).  
 RN [6]  
 RP VARIANTS MPS-I GLN-89.  
 RX MEDLINE-94027086; PubMed-8213840;  
 RA Scott H.S., Litjens T., Nelson P.V., Thompson P.R., Brooks D.A.,  
 RT Hopwood J.J., Morris C.P.;  
 RT "Identification of mutations in the alpha-L-iduronidase gene (IDUA)  
 RT that cause Hurler and Scheie syndromes.";  
 RL Am. J. Hum. Genet. 53:973-986(1993).  
 RN [7]  
 RP VARIANTS MPS-I PRO-366 AND ARG-409.  
 RX MEDLINE-93318833; PubMed-8328452;  
 RA Bach G., Moskowitz S.M., Tieu P.T., Matynia A., Neufeld E.F.;  
 RT "Molecular analysis of Hurler syndrome in Druze and Muslim Arab  
 RT patients in Israel: multiple allelic mutations of the IDUA gene in a  
 RT small geographic area.";  
 RL Am. J. Hum. Genet. 53:330-338(1993).  
 RN [8]  
 RP VARIANTS MPS-I ARG-533.  
 RX MEDLINE-93250829; PubMed-1301941;  
 RA Scott H.S., Litjens T., Nelson P.V., Brooks D.A., Hopwood J.J.,  
 RT Morris C.P.;  
 RT "Alpha-L-iduronidase mutations (Q70X and P533R) associate with a  
 RT severe Hurler phenotype.";  
 RL Hum. Mutat. 1:333-339(1992).  
 RN [9]  
 RP VARIANTS MPS-I D-51; T-75; P-218; P-327; P-489 AND 16-S--A-19 DEL.  
 RX MEDLINE-95038736; PubMed-7951228;  
 RA Bunge S., Kleijer W.J., Steglich C., Beck M., Zuther C., Morris C.P.,  
 RT Schwinger E., Hopwood J.J., Scott H.S., Gal A.;  
 RT "Mucopolysaccharidosis type I: identification of 8 novel mutations  
 RT and determination of the frequency of the two common  
 RT alpha-L-iduronidase mutations (W402X and Q70X) among European  
 RT patients.";  
 RL Hum. Mol. Genet. 3:861-866(1994).  
 RN [10]  
 RP VARIANTS HIS-33.  
 RX MEDLINE-93138632; PubMed-1362562;  
 RA Scott H.S., Litjens T., Hopwood J.J., Morris C.P.;  
 RT "PCR detection of two RFLPs in exon I of the alpha-L-iduronidase  
 RT (IDUA) gene.";  
 RL Hum. Genet. 90:327-327(1992).  
 RN [11]  
 RP VARIANTS THR-361.  
 RX MEDLINE-94061048; PubMed-8242073;  
 RA Scott H.S., Nelson P.V., Litjens T., Hopwood J.J., Morris C.P.;  
 RT "Multiple polymorphisms within the alpha-L-iduronidase gene (IDUA):  
 RT implications for a role in modification of MPS-I disease phenotype.";  
 RL Hum. Mol. Genet. 2:1471-1473(1993).  
 RN [12]  
 RP VARIANTS MPS-I PRO-490; PRO-492 AND LEU-496.  
 RX MEDLINE-96055518; PubMed-7550232;  
 RA Tieu P.T., Bach G., Matynia A., Hwang M., Neufeld E.F.;  
 RT "Four novel mutations underlying mild or intermediate forms of  
 RT alpha-L-iduronidase deficiency (MPS IS and MPS IH/S).";  
 RL Hum. Mutat. 6:55-59(1995).  
 RN [13]  
 RP VARIANTS MPS-I W-89; 349-D-N-350 DEL; H-383; T-504 AND R-626.  
 RX MEDLINE-96055526; PubMed-7550242;  
 RA Bunge S., Kleijer W.J., Steglich C., Beck M., Schwinger E., Gal A.;  
 RT "Mucopolysaccharidosis type I: identification of 13 novel mutations  
 RT of the alpha-L-iduronidase gene.";  
 RL Hum. Mutat. 6:91-94(1995).  
 RN [14]  
 RP VARIANTS MPS-I ARG-388.  
 RX Bartholomew D.W., McClellan J.M.;  
 RA "A novel missense mutation in the human IDUA gene associated with a  
 RT severe Hurler's phenotype.";  
 RL Hum. Mutat. 12:291-291(1998).  
 RN [15]  
 RP CATALYTIC ACTIVITY: Hydrolysis of alpha-L-iduronosidic linkages in  
 RT sulfated dermatan.  
 CC [1-] SUBUNIT: MONOMER (PROBABLE).  
 CC [1-] TISSUE SPECIFICITY: FOUND UBIQUITOUSLY.  
 CC [1-] DISEASE: DEFECTS IN IDUA ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  
 CC TYPE I (MPS-I), A DISEASE CHARACTERIZED BY THE ACCUMULATION OF THE  
 CC GLYCOSAMINOGLYCANS HEPARAN SULFATE AND DERMATAN SULFATE. PATIENTS  
 CC WITH SEVERE MPS-I (HURLER SYNDROME) USUALLY PRESENT WITHIN THE  
 CC FIRST YEAR OF LIFE A COMBINATION OF HEPATOSPLENOMEGALY, SKELETAL  
 CC DEFORMITIES, CORNEAL CLOUDING AND SEVERE MENTAL RETARDATION.  
 CC OBSTRUCTIVE AIRWAYS DISEASE, RESPIRATORY INFECTION AND CARDIAC  
 CC COMPLICATIONS USUALLY RESULTS IN DEATH BEFORE 10 YEARS OF AGE.  
 CC PATIENTS WITH MILD MPS-I (SCHEIE SYNDROME) MAY HAVE LITTLE OR NO  
 CC NEUROLOGICAL INVOLVEMENT, NORMAL STATURE AND LIFE SPAN, BUT ARE  
 CC CHARACTERIZED BY THE DEVELOPMENT OF JOINTS STIFFNESS, MILD  
 CC HEPATOSPLENOMEGALY, AORTIC VALVE DISEASE AND CORNEAL CLOUDING. THE  
 CC INTERMEDIATE FORM (HURLER/SCHEIE) MAY BE CHARACTERIZED BY  
 CC RELATIVELY LITTLE NEUROLOGICAL INVOLVEMENT, BUT MOST OF THE  
 CC SOMATIC SYMPTOMS DESCRIBED FOR SEVERE MPS-I DEVELOP IN THE EARLY  
 CC TO MID-TEENS, CAUSING CONSIDERABLE LOSS OF MOBILITY.  
 CC [1-] SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M74715; AAA81589.1; -;  
 CC EMBL; M95740; AAA51698.1; -;  
 CC EMBL; M95739; AAA51698.1; JOINED.  
 CC Genew; HGNC:5391; IDUA.  
 CC MIM; 252800; -;  
 CC InterPro; IPR000514; Glyco\_hydro\_39.  
 CC Pfam; PF01229; Glyco\_hydro\_39; 1.  
 CC PRINTS; PR00745; GLYDRIASE39.  
 CC PROSITE; PS01027; GLYCOSYL HYDROL\_F39; 1.  
 CC Hydrolase; Glycosidase; Lysosome; Signal; Disease mutation;  
 CC Polymorphism; Mucopolysaccharidosis.  
 CC SIGNAL 1 27 POTENTIAL.  
 CC CHAIN 28 653 ALPHA-L-IDURONIDASE.  
 CC FT ACT\_SITE 182 182 PROTON DONOR (POTENTIAL).  
 CC FT ACT\_SITE 299 299 NUCLEOPHILE (POTENTIAL).  
 CC FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 415 415 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT CARBOHYD 451 451 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC FT VARIANT 16 19 MISSING (IN MPS-I; HURLER).  
 CC FT VARIANT 33 33 Q -> H.  
 CC FT VARIANT 51 51 G -> D (IN MPS-I; HURLER).  
 CC FT VARIANT 75 75 A -> T (IN MPS-I; HURLER).  
 CC FT VARIANT 82 82 H -> P (IN MPS-I; HURLER/SCHEIE).  
 CC FT FTId=VAR\_003350.  
 CC FT FTId=VAR\_003351.  
 CC FT FTId=VAR\_003352.  
 CC FT FTId=VAR\_003353.

FT VARIANT 89 89 R -> Q (IN MPS-I; SCHEIE; IN JAPANESE  
 FT 21% OF ALLELES).  
 FT /FTIG-Var\_003354.  
 FT R -> W (IN MPS-I; SCHEIE).  
 FT /FTIG-Var\_003355.  
 FT R -> Q.  
 FT /FTIG-Var\_003356.  
 FT G -> R.  
 FT /FTIG-Var\_003357.  
 FT L -> P (IN MPS-I; HURLER).  
 FT /FTIG-Var\_003358.  
 FT V -> A.  
 FT /FTIG-Var\_003359.  
 FT D -> Y (IN MPS-I).  
 FT /FTIG-Var\_003360.  
 FT A -> P (IN MPS-I; HURLER).  
 FT /FTIG-Var\_003361.  
 FT D -> N (IN MPS-I; HURLER).  
 FT /FTIG-Var\_003362.  
 FT MISSING (IN MPS-I; HURLER).  
 FT /FTIG-Var\_003363.  
 FT A -> T.  
 Query Match 7.0%; Score 86; DB 1; Length 653;  
 Best Local Similarity 20.7%; Pred. No. 14;  
 Matches 53; Conservative 39; Mismatches 100; Indels 64; Gaps 10;  
 QY 25 ESISPLGKAGQALSHLVAPHRFDKAIYSSVGLPDEE-----AEATVRSQA 70  
 DB 243 DGTNFTCEAGVRLDY-LSLHR--KGARSISILSEQKVVAQIQRLPKFADTYINDE 299  
 QY 71 AASLMGTTETPVLRLQYDRRLTGADRLAAVVGASHLPNTLLVIDAGTATYERVSAAE 130  
 DB 300 ADPL-VGWSLPQWR-----ADVTYAAVMVVKVIAQHONLLANTTSAPPYALLSND 349  
 QY 131 GYL-----GGNISPLCLHRLFRKALHLFTGRPLPLDPSPGISPKIAEYG 172  
 DB 350 NAFLSYHPHPAQRILTARFQVNNTRPPHVQLLRKPVLTAMGLLALLDEEQLWAEVSQAG 409  
 QY 173 SSTEAITAGVI---HGLAGEIDRY-----IDDLHAKGRSAVIL-----TGGDA 214  
 DB 410 TVLDSNHTVGLASAHRRPQGPADAAWRAAVLIYASDDTRAHPNRSVAVTLRGRVPPGCL 469  
 QY 215 NYLARIISGILIHDP 230  
 DB 470 VIVTRYLDNG-ICSPD 484  
 RESULT 11  
 WHX\_BACSU  
 J AMRX\_BACSU STANDARD; PRT; 389 AA.  
 AC P54983;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Amidohydrolase amhX (EC 3.5.1.1-) (Aminoacylase).  
 GN AMHX.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=96313308; PubMed=8768514;  
 RA Kempf B., Bremer E.;  
 RT "A novel amidohydrolase gene from Bacillus subtilis cloning: DNA-  
 sequence analysis and map position of amhX.";  
 RL FEMS Microbiol. Lett. 141:129-137(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97124189; PubMed=8969502;  
 RA Yamane K., Kumano M., Kurita K.;

RT "The 25 degrees-36 degrees region of the Bacillus subtilis  
 chromosome: determination of the sequence of a 146 kb segment and  
 identification of 113 genes.";  
 RL Microbiology 142:3047-3056(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Hajech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones D.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche M., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [4]  
 RP SEQUENCE OF 335-389 FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=95348093; PubMed=7622480;  
 RA Kempf B., Bremer E.;  
 RT "OpuA, an osmotically regulated binding protein-dependent transport  
 system for the osmoprotectant glycine betaine in Bacillus subtilis.";  
 RL J. Biol. Chem. 270:16701-16713(1995).  
 CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M40.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U47860; AAC44168.1; -;  
 DR EMBL; D50453; BAA08935.1; -;  
 DR EMBL; Z99105; CABL2095.1; -;  
 DR EMBL; U17292; -; NOT\_ANNOTATED\_CDS.  
 DR MEROPS; M40.UNN; -;  
 DR SubtilList; BG11789; amhx.  
 DR InterPro; IPR002933; Peptidase\_M20.  
 DR Pfam; PF01546; Peptidase\_M20; 1.  
 KW Hydrolase; Complete proteome.  
 FT CONFLICT 161  
 FT CONFLICT 367 389 R -> A (IN REF. 1 AND 4).  
 FT CSKHFRRRNHLLPLTLNKKAA -> VLKTFQKESLAAA  
 FT NAS (IN REF. 1 AND 4).  
 FT SEQUENCE 389 AA; 42530 MW; 03856E70AA3F5D9F CRC64;  
 Query Match 6.9%; Score 85; DB 1; Length 389;  
 Best Local Similarity 24.7%; Pred. No. 9;

Matches 42; Conservative 23; Mismatches 65; Indels 40; Gaps 8;

QY 22 NSIESISFLPGKAG-----QALSHLVAPHRFDKAIYSSVGLP-----DEEAE 63  
DB 193 NSIEIAAFVHLGLIHDPQHPVTKYKQLQAGESSNIIFPKASFSLDLRAQTNEAME 252  
QY 64 AIV-----RSCAAALMMGT-----ETPVPLRLQ-YDRRTLGADRLAAVVGVAHSLVPNTE 112  
DB 253 ALIAETERACEAAAFAGKIELHKEHSLPANTONKEAEINAEATEIIGAERL---DD 309  
QY 113 LLVIDAGT-----AITYERSVAGIYLGNIISPLG-----HLRFKALHLFTG 154  
DB 310 PLVTTGGEDFHFYAVKVPNKLTMLGLGGLQPLGLHPHMTFDRNAMFTG 359

RESULT 12  
ID GCSP\_MYCTU STANDARD; PRT; 941 AA.  
C Q50601;  
F 01-NOV-1997 (Rel. 35, Created)  
T 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable glycine dehydrogenase [decarboxylating] (EC 1.4.4.2) (Glycine  
DE decarboxylase) (Glycine cleavage system P-protein).  
GN GCVP OR GCVB OR RV1832 OR MT1880 OR MTCY1A1.11C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP STRAIN=H37RV;  
RC MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF  
CC GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE  
CC THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND  
CC THE REMAINING METHYLAMINE MOIETY IS THEN TRANSFERRED TO THE  
CC LIPOLAMIC COFACTOR OF THE H PROTEIN (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: Glycine + lipoylprotein - S-  
CC aminomethylidihydropyridoxalprotein + CO(2).  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:  
CC P, T, L, AND H (BY SIMILARITY).  
CC -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
DR EMBL; Z78020; CAB01470.1; --  
DR EMBL; AE007046; AAK46152.1; --  
DR TIGR; MT1880; --  
DR TubercuList; RV1832; --  
DR InterPro; IPR003437; GDC-P.  
DR Pfam; PF02347; GDC-P; 1.  
DR TIGRFAMs; TIGR00461; gcvp; 1.  
DR Oxidoreductase; Pyridoxal phosphate; Complete proteome.  
KW Oxidoreductase; Pyridoxal phosphate; ALA-RICH.  
FT DOMAIN 437 446  
FT BINDING 692 692 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 941 AA; BFFDE54EDA56B914 CRC64;  
  
Query Match 6.9%; Score 84.5; DB 1; Length 941;  
Best Local Similarity 22.1%; Pred. No. 28;  
Matches 62; Conservative 40; Mismatches 92; Indels 87; Gaps 14;  
  
QY 25 ESTSELPGRAGQA-LSHLVAPHRFDKAIYSSVGLPDEEAEIVRSC-----AAASILMMCTE 79  
DB 547 DAVSLQPNAGSQGEYAGLLAIHEY-----HASRGEPRHDIPLPSAHTGNTNAASALAGMR 602  
QY 80 TPVPLRLQYDRRTLGADRLAAVVGVAHSLVPNTE-----LLVIDAGT 120  
DB 603 -VVVDCHDNGVDVLDLRLAKVGEHAERLSALMTIYTPSTGVYEHDAICAAVHDAGG 660  
QY 121 AITYERSVAGIY-----LGNIS-----PGL-----HLR-FKAL 149  
DB 661 QVYVDGANLNALVGLARPGRGFQVSHLNLHKTFCIPHGSGGPGVAVRAHLAFLPG 720  
QY 150 HLTGRLPLIDPSGISPKIAEYSSTEATAGVIHGLAGEIDRYIDDLHAKGRSAVIL 209  
DB 721 HPFAPEL-----PKGYPVSSAPYGSASILPITWA-----YIRMMGAEGURASLT 765  
QY 210 TGGDANVLARI-----RSGILHPDLV-LIGLNR 239  
DB 766 AITSANYIARRLDEYVPLVTGNGVWAHECILDRLGITKL 806  
  
RESULT 13  
ID PYRD\_MYCTU STANDARD; PRT; 357 AA.  
AC C06236;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)  
DE (DHODase) (DHODase) (DHOD).  
GN PYRD OR RV2139 OR MT2197 OR MTCY270.29C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF  
CC GLYCINE. THE P PROTEIN BINDS THE ALPHA-AMINO GROUP OF GLYCINE  
CC THROUGH ITS PYRIDOXAL PHOSPHATE COFACTOR; CO(2) IS RELEASED AND  
CC THE REMAINING METHYLAMINE MOIETY IS THEN TRANSFERRED TO THE  
CC LIPOLAMIC COFACTOR OF THE H PROTEIN (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: Glycine + lipoylprotein - S-  
CC aminomethylidihydropyridoxalprotein + CO(2).  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:  
CC P, T, L, AND H (BY SIMILARITY).  
CC -!- SIMILARITY: TO OTHER GLYCINE CLEAVAGE SYSTEM P-PROTEIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) -> orotate +  
 CC H(2O)(2).  
 CC -1- COFACTOR: FMN (BY SIMILARITY).  
 CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.  
 CC SUBFAMILY 2.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; Z95388; CAB08654.1; -;  
 EMBL; AE007067; AAK46481.1; -;  
 TIGR; MT2197; -;  
 DR TubercuList; RV2139; -;  
 DR InterPro; IPR001295; DHQ\_dh.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR Pfam; PF01180; DHodehase; 1.  
 DR TIGRFAMS; TIGR01036; PyrD\_sub2; 1.  
 DR PROSITE; PS00911; DHODEHASE.1; 1.  
 DR PROSITE; PS00912; DHODEHASE.2; 1.  
 KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;  
 KW Complete proteome. 294 FMN (POTENTIAL).  
 FT NP\_BIND 286  
 SQ SEQUENCE 357 AA; 399D107DD9B4FCB6 CRC64;

Query Match 6.8%; Score 84; DB 1; Length 357;  
 Best Local Similarity 26.1%; Pred. No. 9.8;  
 Matches 64; Conservative 27; Mismatches 100; Indels 54; Gaps 12;

QY 1 MSFNLIVDQGSACKVAFVRNNSIESIFLPGK-----AGQALSHLVAPHRFDAIYS-- 53  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 117 MGFN---NHGARALAIRLARPEIPGVNIGTKTKTPAGDAVNDYRASARVGPLASYL 173  
 QY 54 -----SVGLPDEEAEIYRSCAAASLMGTETPPLRLQYDRRTLGADRLAAVYGAH 105  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 174 VNVSSNTPGLRDLQAVESRPLISA-VRAETSTPVLVKTAPLSDSDLDLDIADLAVEL 232  
 106 SLYPNTTELLVTDAGTAITYERSVAEGIYL-----GGNISPGHLR-----FKALHFLT 153  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 233 DL-----AGIVATNTVSRDGLTTPGVDRLPGGIGSPPLAQRAVQVLRRLDRV 282  
 QY 154 G-RPLIDPGISPKIAEYSGSTEEATAG---VTHGLAGEI---DRYIDDLHFAKGRSAV 207  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 283 GDRLLAISVGGI-----ETADDAWERITAGASLLQGYTGFTYGGERWAKDIH--EGIARR 335  
 QY 208 ILTGG 212  
 Db : : : : :  
 336 LHDGG 340

## RESULT 14

LU1B\_LYCPN STANDARD; PRT; 612 AA.  
 ID LU1B\_LYCPN  
 AC 004974;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 2-isopropylmalate synthase B (EC 4.1.3.12) (Alpha-isopropylmalate  
 synthase B) (Alpha-IPM synthetase B).  
 GN IPMSB.

OS Lycopersicon pennellii (Tomato).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asterales; easterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=28526;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf;  
 RA Wei T., Maita D., Steffens J.C.;  
 RT "Cloning of two L. pennellii 2-isopropylmalate synthase cDNA and  
 RT their functional expression in yeast";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Catalyzes the condensation of the acetyl group of  
 CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form  
 CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
 CC -1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA ->  
 CC acetyl-CoA + 3-methyl-2-oxobutanoate + H(2O).  
 CC -1- PATHWAY: Leucine biosynthesis; first step.  
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE  
 CC SYNTHASE FAMILY. LEU1A SUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; AF004166; AAB61599.1; -;  
 DR InterPro; IPR002034; AIPM/Hcit\_synth.  
 DR InterPro; IPR000891; HMGL-like.  
 DR Pfam; PF00682; HMGL-like; 1.  
 DR TIGRFAMS; TIGR00973; leuA\_bact; 1.  
 DR PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH.1; 1.  
 DR PROSITE; PS00816; AIPM\_HOMOCIT\_SYNTH.2; FALSE\_NEG.  
 KW Leucine biosynthesis; Lyase.  
 SQ SEQUENCE 612 AA; 66535 MW; 2EC5AEC9039791B CRC64;

Query Match 6.8%; Score 84; DB 1; Length 612;  
 Best Local Similarity 23.3%; Pred. No. 19;  
 Matches 60; Conservative 31; Mismatches 100; Indels 66; Gaps 12;

QY 1 MSFNLIVDQGSACKVAFVRNNSIESIFLPGKAGQA-----LSHLVAPHRFDAIYSSVG 56  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 187 MSRDQVVEKARS--NVAYARSIGCEDVEFSPEDAGSDPEFLYHILG--EVIKAGATTEN 242  
 QY 57 LPDEEAEIYRSCAAASLMGTETP-----VPLRLQYDRRTLGADRLA-AVVGASLTP 109  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 243 IPDTVGYTVPEEFGQLIAKIRANTPGVEDVIISTHCNDLGLSTANTLAGACAGARQLEV 302  
 QY 110 NTELLVIDAGTAITYERSVA-----EGYLVGNIS-----PGLHLR-FK 147  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 303 TINGIGERAGNASLEEVYMAKRCGEVQLGGYTGINTQHILMSSKMVGISGLHVOPHK 362  
 QY 148 ALHLFTGLPLIDPSG-----ISPKIAEYSGSTEEATAGVIHG----- 186  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 363 AI---VGANAFVHESGIHQDGMKHKDITYEISPEDIGLNANESGIVFGKISGVMLCKP 419  
 QY 187 ----LAGEID-RYIDDL 198  
 Db ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 420 KMELGYEIEGKELDDL 436

## RESULT 15

FABG\_ACTAC STANDARD; PRT; 242 AA.  
 ID FABG\_ACTAC  
 AC P70720;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100) (3-ketoacyl-  
 DE acyl carrier protein reductase).



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.051 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIYDQNSACKVAFVR.....ILIHDPVLVLLGLNRLILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1229	100.0	244	23 AAU91168	Pantothenate kinas
2	224	18.2	258	23 AAU91153	Rhodobacter capsul
3	221	18.0	257	23 AAU91174	Pantothenate kinas
4	216	17.6	259	23 ABB47661	Listeria monocytog
5	210.5	17.1	258	23 AAU91172	Pantothenate kinas
6	208	16.9	255	23 AAU91154	Geobacter sulfurre
7	207.5	16.9	262	23 AAU91170	Pantothenate kinas
8	207	16.8	258	22 AAU01243	B. subtilis novel
9	207	16.8	258	23 AAU91149	Bacillus subtilis
10	205.5	16.7	265	23 AAU91151	Streptomyces coeli

11	201.5	16.4	256	23 AAU91175	Pantothenate kinas
12	198	16.1	260	23 AAU91173	Pantothenate kinas
13	197.5	16.1	254	23 AAU91171	Pantothenate kinas
14	195.5	15.9	248	23 AAU91164	Pantothenate kinas
15	184	15.0	219	23 AAU91176	Pantothenate kinas
16	183.5	14.9	241	23 AAU91179	Pantothenate kinas
17	181.5	14.8	242	23 AAU91180	Pantothenate kinas
18	181	14.7	249	23 AAU91178	Pantothenate kinas
19	178	14.5	249	23 AAU91182	Pantothenate kinas
20	175.5	14.3	233	23 AAU91163	Pantothenate kinas
21	174	14.2	212	23 AAU91177	Pantothenate kinas
22	172.5	14.0	273	23 AAU91157	Treponema pallidum
23	167.5	13.6	250	23 AAU91150	Clostridium acetob
24	166	13.5	455	20 AAU38617	Neisseria gonorrh
25	166	13.5	455	21 AAU74908	Neisseria gonorrh
26	166	13.5	460	23 AAU91167	Pantothenate kinas
27	166	13.5	592	20 AAU38618	Neisseria gonorrh
28	166	13.5	592	21 AAU74911	Neisseria gonorrh
29	163	13.3	267	23 AAU91162	Bordetella pertussis
30	162.5	13.2	389	21 AAU74909	Neisseria meningit
31	162	13.2	246	23 AAU91156	Thermotoga maritim
32	161	13.1	592	20 AAU38615	Neisseria meningit
33	161	13.1	592	21 AAU74912	Neisseria meningit
34	161	13.1	592	23 AAU91169	Pantothenate kinas
35	156	12.7	455	21 AAU74910	Neisseria meningit
36	156	12.7	592	20 AAU38616	Neisseria meningit
37	156	12.7	592	21 AAU74913	Neisseria meningit
38	156	12.7	592	23 AAU91166	Pantothenate kinas
39	153	12.4	272	22 AAG81225	Mycobacterium tube
40	153	12.4	272	23 AAU91152	Mycobacterium tube
41	152	12.4	257	23 AAU91160	Synechocystis pant
42	135	11.0	262	23 AAU91158	Borrelia burgdorfe
43	132.5	10.8	229	23 AAU91159	Aquifex aeolicus p
44	131	10.7	262	23 AAU91155	Deinococcus radiop
45	129.5	10.5	223	19 AAU98422	H. pylori GHPO 344

## ALIGNMENTS

RESULT 1  
AAU91168  
ID AAU91168 standard; Protein; 244 AA.  
XX  
AC AAU91168;  
DT  
DT 05-JUN-2002 (first entry)  
XX  
XX Pantothenate kinase (Coax) #6.  
DE  
DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
OS Porphyrcmonas gingivalis.  
XX  
XX WO200216601-A2.  
PN  
PD 28-FEB-2002.  
XX  
XX 24-AUG-2001; 2001WO-US26531.  
XX  
XX 24-AUG-2000; 2000US-227860P.  
PR  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX Yocum RR, Patterson TA;  
PI  
XX WPI; 2002-269358/31.  
DR  
DR N-PSDB; ABK54189.  
XX  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 8; Page 94-95; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.

XX Sequence 244 AA;

Query Match 100.0%; Score 1229; DB 23; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-119;  
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MSFNLIYDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHREDKAIYSSVGLPDE 60  
 Db 1 MSFNLIYDQGNACKVAFVRNNSIESISFLPGKAGQALSHLVAPHREDKAIYSSVGLPDE 60  
 QY 61 EAEAIYRSCAAASLMGMGTETPVLRLOYDRRTLGADRLAAVVGAAHSLYPNTLLVIDAGT 120  
 Db 61 EAEAIYRSCAAASLMGMGTETPVLRLOYDRRTLGADRLAAVVGAAHSLYPNTLLVIDAGT 120  
 QY 121 AIYVERVSAAGIYLGNNISPLGLHFRKALHFTGRPLIDPSPGKIAEYSGSSTEAIT 180  
 Db 121 AIYVERVSAAGIYLGNNISPLGLHFRKALHFTGRPLIDPSPGKIAEYSGSSTEAIT 180  
 QY 181 AGVIHGLAGEIDRYIDDLHAKESRAVILTGCGDANYLARIIRSGILIHPLDLVLGLNRL 240  
 Db 181 AGVIHGLAGEIDRYIDDLHAKESRAVILTGCGDANYLARIIRSGILIHPLDLVLGLNRL 240  
 QY 241 EYNV 244  
 Db 241 EYNV 244

RESULT 2

AAU91153  
 ID AAU91153 standard; Protein; 258 AA.

XX AAU91153;

DT 05-JUN-2002 (first entry)

Rhodobacter capsulatus pantothenate kinase Coax.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Rhodobacter capsulatus.

OS WO200216601-A2.

PN 28-FEB-2002.

PD 24-AUG-2001; 2001WO-US26531.

PF 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

PT Identifying potential antibiotic or antimicrobial agent, comprises

PT  
 XX  
 PS  
 XX

Claim 10; Page 71-72; 128pp; English.

CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.

XX Sequence 258 AA;

Query Match 18.2%; Score 224; DB 23; Length 258;  
 Best Local Similarity 29.4%; Pred. No. 1.1e-14;  
 Matches 72; Conservative 37; Mismatches 86; Indels 50; Gaps 10;  
 QY 18 FVRNNSIESISFLPGKAGQALSHLVAP-----HREFKAIYSSVGLPDEEAEIV 66  
 Db 40 FVWLTLMQLKGLQGRISSEAISSAPRVFVNLVLCNRYFCRPY-VVGKP----- 90  
 QY 67 RSCAAASLMGMGTETPVLRLOYDRRTLGADRLAAVVGAAHSLYPNTLLVIDAGTAITYER 126  
 Db 91 -----GCELPVAPRVD-PGTVTGPDRLVNTVAGYDRH-GGDLIYVDFGTATFDV 138  
 QY 127 VSAGIYLGNNISPLGLHFRKALHFTGRPLIDPSPGKIAEYSGSSTEAITAGV 183  
 Db 139 VAPDGAYIGGVIAAGVNLSEALHMAAALPHVDVTKPGV-----IGTNTVACIQSGV 192  
 QY 184 IHGLAGEIDRYIDDLHAKESRAVILTGCGDANYLARIIRSGI-----LIHPDLVLGLNRL 238  
 Db 193 YWYIGLVGIVQIRMERDRPMKVIATGG-----LASFLDLGFDLFDKVEDDLTMHGLRL 248  
 QY 239 ILEYN 243  
 Db 249 IFDYN 253

RESULT 3

AAU91174  
 ID AAU91174 standard; Protein; 257 AA.

XX AAU91174;

DT 05-JUN-2002 (first entry)

DE Pantothenate kinase (Coax) #12.

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

OS Chlorobium tepidum.

PN WO200216601-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

DR N-PSDB; ABK54195.



XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX  
 PS  
 XX Claim 10; Page 103-104; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 XX }

Sequence 257 AA;

Query Match 18.0%; Score 221; DB 23; Length 257;  
 Best Local Similarity 28.8%; Pred. No. 2.2e-14;  
 Matches 76; Conservative 43; Mismatches 111; Indels 34; Gaps 8;  
 QY 5 LIVDOGSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEAE 64  
 DB 3 LVDVIGNTSTLAFTGDEEPSVSP-----SALFADSSSTMREYFGNMARKHGEPOA 55  
 QY 65 I-VRSCAAASLMGT-----ETPV-----PLRLQY-DRRTLGADRLAAVGAHS 106  
 DB 56 IAICSVVPSATAGSALLSFVLTICCKLRFPRLDYAPPHFGADRLALCAWSRH 115  
 QY 107 LPNTELLVIDAGTAITYERSAEGIYLGNNISPLGLHFRKALHLFTGRLLIDPSPGISP 166  
 DB 116 LFESEKPIAVDIGTAITFDVLDVGNVYRGLIMPIDMAGALHSRTAQLPQV---RIDR 172  
 QY 167 KIAEYGSSTEEATAGVIHGLAGEIDRYID----DLHAKGRSA--VILTGGDANYLARI 220  
 DB 173 PESLLGRSTTECKSGVFWGVKQIGGLVDAIRGDLVRDFGESTVEIVTGGNSRIIYPE 232  
 QY 221 IIRSGILHPDLVLLGLNRLILEYNNV 244  
 DB 233 IGPVSVIDELAVLRGSDLLLRNMN 256

RESULT 4  
 BB47661  
 D ABB47661 standard; Protein; 259 AA.  
 AC ABB47661;  
 XX  
 XX

DT 05-FEB-2002 (first entry)  
 DE Listeria monocytogenes protein #365.  
 XX  
 XX

DE Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 XX

OS Listeria monocytogenes.  
 XX  
 XX

PN WO200177335-A2.  
 XX  
 XX

PD 18-OCT-2001.  
 XX  
 XX

PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 XX

PR 11-APR-2000; 2000FR-0004629.  
 XX  
 XX

PA (INSP ) INST PASTEUR.  
 XX  
 XX

PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Iopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 XX WPI; 2002-010914/01.

XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 XX  
 XX

PS Claim 5; SEQ ID No 366; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by *L.*  
 CC monocytogenes and related organisms.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 259 AA;

Query Match 17.6%; Score 216; DB 23; Length 259;  
 Best Local Similarity 25.4%; Pred. No. 7.2e-14;  
 Matches 69; Conservative 43; Mismatches 106; Indels 54; Gaps 9;  
 QY 5 LIVDOGSACKVAFVRNNSIESISFLPGKAGQALSH---LVAPHRFDKAI-----YSS 54  
 DB 3 LVDVIGNTCTVGVYKQL-----LKHWRMTDRHRTSDELGMTVLNFFSY 49

QY 55 VGLPDEAEAIVRSCAAASLMGTET-----PVLRLQYDR-RTLGD 96

DB 50 ANLTPSDIQIGIISVVPPIMHAMETCMVYFNIRPLIVGPGIKTGLNKLVDNPREIGSD 109

QY 97 RLAAVVGASHLYPNTELLVIDAGTAITYERSAEGIYLGNNISPLGLHFRKALHLFTGR 156

DB 110 RIVNAVAASEY-GTPVIVDFGATTCYIDESGVYOGGATAPGIMISTEALYNRAKL 168

QY 157 PLIDPSGISPKIAE-----YGSSTEEATAGVIHGLAGEIDRYIDDLHAKGRSAVIL-TG 211

DB 169 PRVD-----IAESSQIIGKSTVSSMQAGIFGVGCEGIIAEMKKQSNASPPVAVTG 221

QY 212 GDANYLARIIRSGILHPDLVLLGLNRLILEYNNV 243

DB 222 GLARMIKSSAVDILDPLFLTKGLELLYRN 253

RESULT 5  
 AAU91172  
 ID AAU91172 standard; Protein; 258 AA.  
 XX  
 XX

AC AAU91172;  
 XX  
 XX

DT 05-JUN-2002 (first entry)  
 XX  
 XX

DE Pantothenate kinase (Coax) #10.  
 XX  
 XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX



```

PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54191.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 8; Page 98-99; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 262 AA;
Query Match 16.9%; Score 207.5; DB 23; Length 262;
Best Local Similarity 30.8%; Pred. No. 5,5e-13;
Matches 72; Conservative 44; Mismatches 73; Indels 45; Gaps 14;
QY 25 ESISFLPKAGQALSHLVAPHRFDKAIYSSVGLPDEAEAIYRSC-----AAASLMGTET 80
DB 50 EGLSPEDYK-GIIYSSVPPIMF-----ALERMCEKFKIKPLVVGPI 92
QY 81 PVPLRLQYDR-RTLGADRLA-AVCGAHSLYPNTELLVIDAGTAIYERVSAEGIYLGNI 138
DB 93 KTGLNIKENPREVGADRIYNAVAGIH-LY-GSLIIVDFGTATTYCYINEEKHYMGVI 150
QY 139 SPGLHLRFKALHFTGRPLID---PSGISPKIAEYGSSTEAITAGVIHGLAGEIDRYI 195
DB 151 TPGIMISAEALYSRAAKLPRIETKPSV-----VGKNTYSAMQSGILYGVGVQEGIV 204
QY 196 DDL--HAKEGRSAVILTGADNYLARI--RSGIL--IHPDLVLLGLNRILEYN 243
DB 205 KRMKEARQ-EPKVIATGG----LAKLISEESNVIDVDPFLTKGLYMLYERN 253
RESULT 8
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coax.
XX
KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.

```

```

PR 24-AUG-2000; 2000US-0227860.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS.
XX
PI Yocum RR, Patterson TA, Hermann T, Pero JG;
XX
DR WPI; 2001-218644/22.
DR N-PSDB; AAS00984.
XX
PT New recombinant microorganism which overexpress a Bacillus subtilis
PT pantothenate biosynthetic enzyme, useful for the high yield production
PT of panto-compounds such as pantothenate and pantoate
XX
PS Example 14; Fig 23; 292pp; English.
XX
CC The sequence represents a novel B. subtilis pantothenate kinase (encoded
CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.
CC Pantothenate, also known as vitamin B5, is used as a nutritional
CC supplement in mammals and humans. The invention concerns methods of
CC producing recombinant microorganisms overexpressing at least one Bacillus
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
CC of producing them are useful for producing a panto-compound such as
CC pantothenate or pantoate, which is a nutritional requirement for
CC livestock and humans. The methods are also useful for the identification
CC of pantothenate kinase modulators. Panto-compounds are produced at a
CC significantly higher yield than prior art methods and can be produced
CC independent of the need to feed precursors which decreases expense.
XX
SQ Sequence 258 AA;
Query Match 16.8%; Score 207; DB 22; Length 258;
Best Local Similarity 26.4%; Pred. No. 6.1e-13;
Matches 71; Conservative 49; Mismatches 101; Indels 48; Gaps 10;
QY 5 LIVDOGSACKVAFVRNNSIESISFLPKAGQALSHLVAPHRFD-----KAIYSSVGL 57
DB 3 LIVDVGNT-----NTVLGV-YHDGKLEYHWRIETSRKTEDEFGMLIRSLFDHSL 52
QY 58 PDEAEAIYRSCAAASLM-----MGTEPPVPLRLQYDR-RTLGADRLA 99
DB 53 MFEQIDGIITISSVPPIMFALERMCYKFIHPIQVPGMKTGILNIKNPNKEVGADRI 112
QY 100 AVCGAHSLYPNTELLVIDAGTAIYERVSAEGIYLGNIISPLGLHLRFKALHFTGRPLI 159
DB 113 NAVAAILHLYGN-PLIYVDFGTATTYCYIDENKQYMGGAIAPIGTISTEALYSRAAKLPRI 171
QY 160 D---PSGISPKIAEYGSSTEAITAGVIHGLAGEIDRYIDDL--HAKEGRSAVILTGDA 214
DB 172 EITRPDI-----IGKNTYSAMQSGILFYGVQGVGIVKRMKQADLK-VIATGGLA 224
QY 215 NYLARIIRSGILHHPDLVLLGLNRILEYN 243
DB 225 PLIANESDCIDIVDPFLTKGLGLEIYERN 253
RESULT 9
AAU91149
ID AAU91149 standard; Protein; 258 AA.
XX
AC AAU91149;
XX
DT 05-JUN-2002 (first entry)
XX
DE Bacillus subtilis pantothenate kinase Coax.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
PN WO200216501-A2.
XX
PD 28-FEB-2002.

```



XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX N-PSDB; ABK541196.  
 PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 6; Page 105; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 256 AA;  
 SQ  
 Query Match 16.4%; Score 201.5; DB 23; Length 256;  
 Best Local Similarity 27.0%; Pred. No. 2.2e-12;  
 Matches 75; Conservative 44; Mismatches 94; Indels 65; Gaps 11;  
 QY 5 LIVDQNSACKVAFVRNNSI-----ESISFLPGKAGQALSHLVAPHFRDKAIYSS 54  
 Db 3 LVFDVGNFMVLIYKDKLVNWKIKDREKTS---DEYGLISNLT-----FD---YDN 51  
 QY 55 VGLPDEAEAI-----VRSAAASLMMGTETVPVLRLOQYDR-RTLGADRL 98  
 Db 52 VNISDIDDVLISSVVPNMVHSLNFCIKYCKQPLIVGPIKTLGNIKYDNPKQVGADRI 111  
 QY 99 AAVGASHLYPNTLLVIDAGTAITYYRSAGIYLGNNISPLHLRFKALHLFTGRPL 158  
 Db 112 VNAVAGIEKYGPSILV-DFGTATTCFAISEKGEYLGSTIAPGKISSEALFQSASKLPR 170  
 QY 159 IDPSGISPKIAYG---SSTEEATAGVIHGLAGEIDRYI-----DDLHAKGRS 205  
 Db 171 VE-----LAKPGMTCKSTVSAMQSGIYGVGLVDKIISIMKKELNCDV----- 216  
 QY 206 AVILTGGDANYLIARIIRGILIHDPDLVLLGLNRLILEYN 243  
 Db 217 KVIATGGLAKLIASETKSIDYVDGFLTEGLRIIYEKN 254  
 RESULT 12  
 AAU91173  
 ID AAU91173 standard; Protein; 260 AA.  
 XX AAU91173;  
 AC AAU91173;  
 DT 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #11.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 KW Caulobacter crescentus.  
 OS WO200216601-A2.  
 PN 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX

PR 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX N-PSDB; ABK541194.  
 PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 102-103; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 260 AA;  
 SQ  
 Query Match 16.1%; Score 198; DB 23; Length 260;  
 Best Local Similarity 33.7%; Pred. No. 5.3e-12;  
 Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;  
 QY 74 LMGTETPVPLRLOQYDRRT-LGADRLAAVVGASHLYPNTLLVIDAGTAITYYRSAGI 132  
 Db 86 LVIGENAKLIGIDVRIEKPSSEAGADRLVNAIGAAVYGP-PLWVIDSGTATTFDVAADGA 144  
 QY 133 YLGGNISPLGLHLRFKALHLFTGRPLI---DPSGISPKIAYEGSSTEEATAGVIHGLAG 189  
 Db 145 FEGGIAPGINLSQALHEAAKLPRIAIQRPAG--NRIV--GTDVTSAMQSGVFWGYIS 200  
 QY 190 EIDRYIDDLHAKGR-SAVLTGGDANYLIARIIRGILIHDPDLVLLGLNRLILEYN 243  
 Db 201 LIEGLVARIKAEKGEPTMTVIATGCVASLFEAGTSDIDHFDSDLTIRGLLEIYRN 255  
 RESULT 13  
 AAU91171  
 ID AAU91171 standard; Protein; 254 AA.  
 XX AAU91171;  
 AC AAU91171;  
 DT 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #9.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 KW Bacillus halodurans.  
 OS WO200216601-A2.  
 PN 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX

XX DR WPI: 2002-269358/31.  
 DR N-PSDB: ABK54192.  
 PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX  
 PS Claim 10; Page 100; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 3 Sequence 254 AA;  
 Query Match 16.1%; Score 197.5; DB 23; Length 254;  
 Best Local Similarity 27.9%; Pred. No. 5.7e-12;  
 Matches 60; Conservative 44; Mismatches 88; Indels 23; Gaps 6;  
 QY 49 KAIYSSVGLPDEAEAIIVSCAAASLMGTE-----TPV-----PLRLQYDR- 90  
 DB 44 RSLFDHAGLQFODIDGIVISSVPPMFSLQMKCKYFHTPTMIIGPKITGLNIKYDNP 103  
 QY 91 RTIGADRLAAVVGHSIYPNTELLVIDAGTAITVERVSAEGIYLGGNISPGHLHFRKALH 150  
 DB 104 KEVGADRIVNAVAIELY-GYPATVDFGTATTCLINEKKQYAGGVTPGIMISTEALY 162  
 QY 151 LFTGRLPLDPSGTSPIAEVGSSTERAITAGVIHGLAGEIDRVIDDHLAK-EGRSVIL 209  
 DB 163 HRASKLPRIE---IAKPKQVGVNTIDMSQGIYGVSDGVVKKMAQAESEPKVIA 219  
 QY 210 TGGDANYLARIIRSGILIHPLDVLVLLNRLILEYNY 244  
 DB 220 TGGGLAKLGTESEFIDVDSFLLTKLQLIYKKNV 254

RESULT 14  
 AAU91164  
 ID AAU91164 standard; Protein; 248 AA.  
 AC AAU91164;

05-JUN-2002 (first entry)  
 Pantothenate kinase (Coax) #2.  
 Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 Pseudomonas aeruginosa.  
 WO200216601-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US26531.  
 24-AUG-2000; 2000US-227860P.  
 20-MAR-2001; 2001US-0813453.  
 (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 Yocum RR, Patterson TA;  
 WPI: 2002-269358/31.

XX PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX  
 PS Claim 6; Page 83-84; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 3 Sequence 248 AA;  
 Query Match 15.9%; Score 195.5; DB 23; Length 248;  
 Best Local Similarity 30.4%; Pred. No. 8.9e-12;  
 Matches 80; Conservative 33; Mismatches 95; Indels 55; Gaps 12;  
 QY 5 LIVDOGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEAEA 64  
 DB 3 LELDCGNSLIKRWVI-EGARSAVA---GGLAESDDALV-----EQLTSQQALP----- 46  
 QY 65 IVRSCAAASLMGTE-----PVP-----LRLQY-DRRTILGADRLAA 100  
 DB 47 -VRACRLVSVRSEQTSOLVARLEQLFPVSALVASSGKQAGVRNGILDYQRLGLDRWLA 105  
 QY 101 VVGHSIYPNTELLVIDAGTAITVERVSAEGIYLGGNISPGHLHFRKALHFTGRPLID 160  
 DB 106 LVAAHHL-AKRACLVIDLGTAVTSDLVADGVHLGGYICPGMTLMRSQRLTRRRIRYDD 164  
 QY 161 P-----SGISPKTAEGSSTEEAITAGVIHGLAGEI-DRYIDDLHAKEGRSVILTGGD 213  
 DB 165 AEARRALASLP-----QOATAEAYERGCLLMRGFVREQYAMACELLGPDCEIFUTGGD 219  
 QY 214 ANYLARIIRSGILIHPLDVLVLLGL 236  
 DB 220 AE-LVRDELAGARIMPDLVFVGL 241

RESULT 15  
 AAU91176  
 ID AAU91176 standard; Protein; 219 AA.  
 AC AAU91176;  
 05-JUN-2002 (first entry)  
 Pantothenate kinase (Coax) #14.  
 Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 Dehalococcoides ethenogenes.  
 WO200216601-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US26531.  
 24-AUG-2000; 2000US-227860P.  
 20-MAR-2001; 2001US-0813453.  
 (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 Yocum RR, Patterson TA;  
 WPI: 2002-269358/31.

Search completed: June 24, 2003, 21:46:29  
Job time : 26.051 secs.

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.8997 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-41  
Perfect score: 1229  
Sequence: 1 MSFNLIYDQNSACKVAFVR.....ILIHDPDLVLLGLNRILEYNV 244

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 73:\*

1: Pirl:\*

2: Pirl:\*

3: Pirl:\*

4: Pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216	17.6	259	2 AF1102	conserved hypot
2	214	17.4	259	2 AF1464	conserved hypot
3	208	16.9	273	2 E97293	probable transcrip
4	205.5	16.7	265	2 T36391	hypothetical prote
5	198	16.1	261	2 B87489	transcription acti
6	197.5	16.1	254	2 F83660	hypothetical prote
7	195.5	15.9	248	2 H83111	hypothetical prote
8	183	14.9	276	2 A12292	hypothetical prote
9	181.5	14.8	242	2 A82637	conserved hypot
10	175.5	14.3	233	2 S66100	conserved hypot
11	172.5	14.0	273	2 D71326	conserved hypot
12	163	13.3	267	2 D140327	baif protein - Bord
13	162	13.2	246	2 D72320	conserved hypot
14	161	13.1	592	2 B81009	BirA protein/Bvg a
15	157.5	12.8	274	2 H86937	conserved hypot
16	156	12.7	592	2 H20465	probable biotin-la
17	153	12.4	272	2 A70955	hypothetical prote
18	152	12.4	257	2 S75559	hypothetical prote
19	135	11.0	262	2 F70165	conserved hypot
20	134	10.9	223	2 G71887	hypothetical prote
21	132.5	10.8	229	2 E70465	hypothetical prote
22	131	10.7	262	2 E75516	conserved hypot
23	129.5	10.5	223	2 F64627	hypothetical prote
24	126.5	10.3	209	2 H81382	hypothetical prote
25	113	9.2	450	2 C95348	HemN coproporphyr
26	105.5	8.6	456	2 JC4089	3alpha,7alpha,12al
27	96.5	7.9	385	2 G87340	aminotransferase,
28	96	7.8	350	2 A75600	conserved hypot
29	95.5	7.8	240	2 G91045	hypothetical prote

30 95.5 7.8 240 2 C85890 hypothetical prote  
31 95.5 7.8 240 2 H65022 hypothetical prote  
32 95.5 7.8 468 1 S19723 dihydroliipoamide d  
33 95 7.7 332 2 G71297 probable fructose-  
34 94.5 7.7 534 2 A98326 ketoglutarate semi  
35 94.5 7.7 534 2 A82957 dehydrogenase Atu3  
36 94 7.6 298 2 F84224 hypothetical prote  
37 93 7.6 323 2 H86785 L-lactate dehydrog  
38 93 7.6 559 2 C87307 conserved hypotet  
39 91.5 7.4 342 2 AB3169 protoporphylin IX  
40 90.5 7.4 758 2 T02925 probable polyketid  
41 90 7.3 1168 2 E70726 L-deoxy-D-xylulose  
42 89.5 7.3 735 2 T08140 glycerate kinase  
43 89 7.2 382 2 AG0567 dihydroliipoamide d  
44 89 7.2 467 2 AD3270 probable ferredoxi  
45 88.5 7.2 409 2 H95911

ALIGNMENTS

RESULT 1

AF1102

conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain E  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AF1102  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AF1102

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-259 <GLA>

A;Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl61409586; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo0221

C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	17.6%	Score 216;	DB 2;	Length 259;
Best Local Similarity	25.4%	Pred. No. 1.7e-10;		
Matches	69;	Conservative	43;	Mismatches 106;
				Indels 54;
				Gaps 9;
QY	5	LIVDQGSACKVAFVRNNSTESISFLPGKAGQALSH---	LVAPHRFDKAI-----YSS	54
DB	3	LVIDVGNTCTGVYEKOKL-----	LKHWRMTTDRHRTSDELGMTVLNFFSY	49
QY	55	VGLPDEAEAIIVSCAAASLMGTET-----	PVPLRLQYDR-RTLGAD	96
DB	50	ANLTPSDIQGIITSSVPPIMHMETMCRVFNTRPLIVGPIKTLGNLKNVDNPREIGSD	109	
QY	97	RLAAVGAHSLYPTNTELLIDAGTAIVYERSVSAEGLVGGNISPGLHRLFKALHFTGLR	156	
DB	110	RYNNAVAASEY-GTPVIVDFGTATTCYIDESGVQGGAIAGIMISTEALYNRAKL	168	
QY	157	PLIDPSGISPKIAB-----YGSSTEEATAGVIHGLAGEIDRYIDLHAKERSAVIL-TG	211	
DB	169	PRVD-----IAESSQIICKSTVSSMQAGIFYGVGOCEGIIAEMKKSNASPVVWATG	221	
QY	212	GDANYLARIIRSGILIHDPDLVLLGLNRILEYN	243	
DB	222	GLARMITEKSAVDILDFFLTGLGLELLYREN	253	

RESULT 2

AF1464

conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Cllp112

```

Db      97  LWVGPGIKTGINIKYDNPKEVGADRIYVNAVAHEIYKRS-LIIIDFGTATTFCAVRENGD 155

Qy      133  YLGNISPGHLRKFALHLFTGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEID 192
      ||||| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||:
Db      156  YLGAICPGIKVSSEALFEKAKLPVEL--IKPAYA-ICKNTLISSIOSGIVGYIGQV- 211
      ||||| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||:

Qy      193  RYI-----DLHAK-EGRSAVILTGGDANYLARIIRSGILIHDPVLVLGLNRILEYN 243
      ||| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||: : :|| :||:
Db      212  RYIVERMKEELOEGEREPLVATGGLAKLISEEAKNVNDVINPFTLEGLRIIYEKN 268

RESULT 4
T36391
Hypothetical protein SCB94.31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
Submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265 <OLI>
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A:Experimental source: strain A3(2)
C:Genetics:
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 16.7%; Score 205.5; DB 2; Length 265;
Best Local Similarity 33.5%; Pred. No. 1.3e-09;
Matches 59; Conservative 36; Mismatches 64; Indels 17; Gaps 7;

Qy      77  GTETPVPLRQYDRRTIGADRLAAVGAHSLYPNTELLVIDAGTAITYYRVVSAGIYLG 136
      | :||| : : :||| : : :||| : : :||| : : :||| : : :||| : : :|||
Db      96  GVTGTGVPILTDHPKE-VGADRIINAAVAVELY-GGPAIVVDGFTATTFDAVSARGEYIG 153
      | :||| : : :||| : : :||| : : :||| : : :||| : : :||| : : :|||

Qy      137  NISPGHLRKFALHLFTGRPLIDPSGISPKIAEYGSSTEEAITAGVIHGLAGEID--- 192
      | :||| : :||| : :||| : :||| : :||| : :||| : :||| : :|||

Db      154  VIAPGTEISYEALGVKAQLRKIE---VARPSRVIGKNTVEAMQSGIVYGFAGQVDGVN 210

Qy      193  RYIDDLHAKEGRSAVILTGGDANYLARIIR--RSGIL--IHPDVLVLGLNRILEYNV 244

Db      211  RMARELADDDPDDVTVIATGG----LAPWLGVSSVIDEHEPWLTLAAGLRLYVERN 262

```

RESULT 5  
 B87489  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C:Accession: B87489  
 R:Rieman, W.C.; Feidlyblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; MUID:21173698; PMID:11259647  
 A:Accession: B87489  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-261 <STO>  
 A:Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:CN00148  
 C:Genetics:  
 A:Gene: CC1935  
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
  
 Query Match 16.1%; Score 198; DB 2; Length 261;  
 Best Local Similarity 33.7%; Pred. No. 5.4e-09;  
 Matches 59; Conservative 34; Mismatches 72; Indels 10; Gaps 6;  
 QY 74 LMMGTETPPVLRQYDRRT-LGADRLLAAVVGAAHSLYPNTELLVIDAGTAITYRVSAEGI 132

[illegible]

## RESULT 6

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83660  
C:Update: 15-Jun-2001  
C:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: F83660  
A:Status: preliminary  
A:Keywords: DNA  
A:Molecule type: DNA  
A:Residues: 1-254 <STO>  
A:Cross-references: GB:AP001507; GB:BA000004; MID:g10172612; PIDN:BA803805.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	16.1%; Score 197.5; DB 2;	Length 254;
Best Local Similarity	27.9%; Pred. No. 5.7e-09;	
Matches	60; Conservative 44; Mismatches 88;	Indels 23; Gaps 6
Qy	49 KAIYSSVGLPDEEAIVRSCAAASLMGTG-----TPV-----PLRLQYDR-	90
Db	44 RSLFDHAGLPQDIDIGVISSVVPMMFSLQMCKKYPHVTPMIIGPGIKTGLNKIYNP	103
Qy	91 RTLGAURLAUVVGASHLPTNTELLVIDAGTAITTVRSVAEGIIYLGNGTSPGULHLRFKAL	150
Db	104 KEVGADRIVNVAIAELY-GYPATVDFGTATTCTCLINEKKQYAGGVAPAGIMIGISTEALY	162
Qy	151 LFTGRLLPIDPSGISPKIAEYSGSTEEATAITAGVTIHGLAGEIDRYIDDLHAK-EGRSAVIL	209
Db	163 HRASKLPRIE---LAKPQQVGTWTFIDSMQSGIFGYVSQVDVVVKRMKAASEPKVIA	219
Qy	210 TGGDANYLARIRSGILITHPLVLGLNLRILEYNV	244
Db	220 TGGLAKLIGTYESETIDVDSPLTKGLQLIYKKNV	254

## RESULTS. 7

H83071  
H83111  
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83111  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: H83111  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-248 <STO>  
A:Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4279

Query Match	15.9%;	Score	195.5;	DB	2;	Length	248;
Best Local Similarity	30.4%;	Pred.	No. 8.1e-09;				
Matches	80;	Conservative	33;	Mismatches	95;	Indels	55;
Gaps	12;						
QY	5	LIVDOGNSACKYAFVFRNNSSIESISFLPGKAGQASHLVAPHRFDKAIYSSVGLPDEEAEA	64				
Db	3	LELDGNSLIKWRI-EGAARSVA--GGLAESDDALV-----EQLTSQALP-----	46				
QY	65	IYRSCAAASLMMGTET-----PVP-----LRLOY-DERTTGAORLAA	100				
Db	47	-VRACRLVSVRSEQTSQLVARLEQFPYVALVASSGKQLAGVRNGYLDVORGLDRKLA	105				
QY	101	VYGASHLPNTPELLVIDAGTAITYERVSAGIYLGGNISPLGHLRFKALHFTGRPLID	160				
Db	106	LVAHHL-AKKACLVIDLTGTVTSDLVAADGVLHGVCYICPGMTLMRSQLTRRIRYDD	164				
QY	161	P-----SGISPKIAEYGSSTEEATAGVIHGLAGEI-DRY IDDLHAKEGRSVAILTGGD	213				
Db	165	AEARRALASLP-----GQATEAVERGCULMLRGFVREQYAMACELLGPDCEIFLTGGD	219				
QY	214	ANYLARIIRSGILIHDPDLVLLGL	236				
Db	220	AE-LYRDELAGARIMPDILVFVGL	241				

8 JUL 1958

hypothetical protein alr3896 [Imported] - Nostoc sp. (strain PCC 7120)  
A:Accession: A12292  
C:Species: Nostoc sp.  
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: A12292  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iino,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Terashima,  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MID:21595285; PMID:11759840  
A:Accession: A12292  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-276 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17l33030; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3896

	Query Match	14.9%;	Score 183;	DB 2;	Length 276;
	Best Local Similarity	31.3%;	Pred. No. le-07;		
	Matches	52;	Conservative 33;	Pred. Matches 71;	Indels 10; Gaps 4;
QY	82	VPLRLOYDRTIGADRLAAVGAHSILYPNTELLVIDAGTAITTYERYVSAGGIYIGGNISPG	141		
DB	110	IPLNNIYP--TGIDRALALWAGMSW-GFPVLIVIDAGTALTTFADGGKNLYGGALLPG	166		
QY	142	LHLRFKALHLFGRLPLIDPSGIPKIAEYSGSTEEATAGVYHGLAGEIDRIDDDLHAK	201		
DB	167	VLQPFASLQQTGQLPQVEMEAIKSLPPRFALNTEAIOGSVIYTLIAGNRDFTTEWLSL	226		
QY	202	EGRSAVILTGGD-----ANYLARI---IRSGILHPDLVLGLNRLI	240		
DB	227	FPDGKVAITKGGRIILLNLYQALYLPDLAARLIVEPNLIWFQMOTIV	272		

RESIST O

conserved hypothetical protein xp1795 [imported] - *Xylella fastidiosa* (strain 9a)  
 AB2637  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: AB2637  
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347  
A>Note: for a complete list of authors see reference number A59328 below  
A:Accession: A82637  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-242 <SIM>  
A:Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:GN001  
A:Experimental source: strain 9a3c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facinanci, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: de Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1795

Query Match 14.8%; Score 181.5; DB 2; Length 242;  
Best Local Similarity 28.0%; Pred. No. 1.1e-07;  
Matches 72; Conservative 39; Mismatches 91; Indels 55; Gaps 12;  
QY 6 IVDOGNSACKVAFVRNNSIESISFLP-----GKAGQALSHLVAPHREFDK 49  
DB 5 LFDLGNRFKASLRGCVGTPVTLPTETMDAFALQELPGRVAY-LASVAAP----- 58  
QY 50 ATYS-----SVGLPDEEAIVRSCAAASLMMGTETPLRLQYDR-RTLGADRLAAVV 102  
DB 59 AITTHVLEVLKTHFEQVQVAATVAACAG-----VRAYAHPEFEGVDRLALL 106  
QY 103 GAHSLYPNTELIVDAGTAITVERVSAEGIYLGNNISPLHLRFRKALHLFTGRPLDPS 162  
DB 107 GS---YGENVLVGVGTALTDLAANGCHLGRISASPTLMRQALHARAEOPL---S 160  
QY 163 GISPKIAEYGSSTEAITAGVIHGLAGEIDRVIDDLHAKGRSAVILT--GGDANYLARI 220  
DB 161 G--GNYLEPAETEDALVSGCGAAVALIERSLYEAHQRDOSVRLLLHGGGVASLPLWL 218  
QY 221 IRSGILIH-PDLVLLGL 236  
DB 219 --GDVVRHTLVLDGL 232  
RESULT 10  
J66100  
Conserved hypothetical protein yacB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000  
C:Accession: S66100; E69740  
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994  
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom  
A:Reference number: S65967; MUID:96051385; PMID:7584024  
A:Accession: S66100  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <OGA>  
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:dl005847; PID:g467459  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma  
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, I.M.; Portet  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scan  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Teramoto, T.; Togononi, A.; Tosato, V.; Uchly  
T.; Winters, P.; Wipat, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Ychlid  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.; Yata, K.; Ychlid  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69740  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <KUN>  
A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el182  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yacB  
A:Start codon: TTG  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 14.3%; Score 175.5; DB 2; Length 233;  
Best Local Similarity 25.0%; Pred. No. 3.4e-07;  
Matches 54; Conservative 43; Mismatches 74; Indels 45; Gaps 8;  
QY 5 LIVDOGNSACKVAFVRNNSIESISFLPCKAGQALSHLVAPHREFD-----KATYSSVGL 57  
DB 3 LVIDGNT-----NTVLGV-YHDGKLEYHWRITSRHKTEDEFGMILRSLFDHSGL 52  
QY 58 PDEEAIVRSCAAASLMMGTETPLRLQYDR-RTLGADRLA 99  
DB 53 MFEQIDGIISSVPPIMFALERMCTYFHEPOIVGPMGTGNIKYNKVDNKEVGADRIV 112  
QY 100 AVVGAHSLYPNTELLVIDAGTAITVERVSAEGIYLGNNISPLHLRFRKALHLFTGRPLI 159  
DB 113 NAVAATHLYGN-PLIVDFCTATTTCYIDENKQTMGAIAPIGTISTEALYSRAAKLPRI 171  
QY 160 D---PSGISPKIAEYGSSTEAITAGVIHGLAGEID 192  
DB 172 EITRPDNI-----IGKNTVSAQSGILFGYGVQVE 201  
RESULT 11  
D71326  
Conserved hypothetical protein TP0431 - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D71326  
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.;  
rson, J.; Khaliq, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.;  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770; PMID:9665876  
A:Accession: D71326  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-273 <COL>  
A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g33  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0431  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 14.0%; Score 172.5; DB 2; Length 273;  
Best Local Similarity 28.9%; Pred. No. 7.3e-07;  
Matches 55; Conservative 33; Mismatches 77; Indels 25; Gaps 8;  
QY 62 ABAIVRSCAAASLMMGTETPLRLQYDRRTLGADRLAAVGAHSLYPNTEL 113  
DB 77 ADAVAQISGVQPVVFGFWAYEHLPLVRIPEVRAE-----IGTDLVANAVAAY-VHRSAC 130  
QY 114 LVIDAGTAITVERVSAEGIYLGNNISPLHLRFRKALHLFTGRPLIDPSGISPKIAEYGS 173

Db 131 VVDCGTRALFTAVDGTGLIOGVAIAPGLRTAVQSLHTGTAQLPLV-PLALPDSV--L GK 187

Qy 174 STEBATAGVTHGLAGEIDRYIDDLHAKEG-RSAVILTGGDANYLARIIRSGILIH P--- 229

Db 188 DTHAVQAGVYRGTFLEIRAMIAQCQKELGCRCAAVITGG----LSRLFSSEVDFPPIDA 243

Qy 230 DLVLLGLNRI 239

Db 244 OLTLGLAHI 253

RESULT 12  
I40327

baf protein - Bordetella pertussis  
C:Species: Bordetella pertussis  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999  
C:Accession: I40327; S70669  
R:DeShazer, D.; Wood, G.E.; Friedman, R.L.  
Bacteriol. 177, 3801-3807, 1995  
A:Title: Identification of a Bordetella pertussis regulatory factor required for transcr  
A:Reference number: I40327; MUID:95325323; PMID:7601846  
A:Accession: I40327  
A:A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-267 <RES>  
A:Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229  
R:Allen, A.; Maskell, D.  
Mol. Microbiol. 19, 37-52, 1996  
A:Title: The identification, cloning and mutagenesis of a genetic locus required for lip  
A:Reference number: S70669; MUID:96419162; PMID:8821935  
A:Accession: S70669  
A:A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 239-267 <ALL>  
A:Cross-references: EMBL:X90711; NID:g9929967; PIDN:CAA62242.1; PID:g9929968  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
C:Genetics:  
A:Gene: baf

```

Query Match      13.3%; Score 163; DB 2; Length 267;
Best Local Similarity 30.8%; Pred. No. 4.3e-06;
Matches 66; Conservative 28; Mismatches 78; Indels 42; Gaps 10;

QY 5 LIYDQNSACKVAFVRNNSIESIFLPGKAGQALSHLVAPHRPDKAIYSVG-----LPD 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 ILIDSGNSRLKVGW-----FDP-DAQA-AREPAPVAFDNLDLALGRWLATLPR 50

QY 60 EEEAEIVRSCA-----AASLMG-----TETVPVLRQYDRT---LGADRLAAV 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 RPORALGVNVAGLARGEALIAATIRAGSGDIRWTLRAQPLAMGLRNGYRNPDLGADRWACM 110

QY 102 VGAHSLYPNTE--LLVIDAGTATTYRVSAGYILGCGNISPLGLHLRFKALHLFTGRPLI 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 VGVLARQPSVHPPELLVASFGTATLTDIGPDNVPFGGLIILPGPAMRGALAYGTAHPLA 170

QY 160 DPSGISPKIAEYGSSTEEAITAGVIHGLAGEIDR 193

Db 171 D--GL---VADYPIDTHOALASGIAAAGCAIYR 199

```

RESULT 13  
D72320

D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: D72320  
R;Neison, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: D72320

A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-246 <ARN>  
A; Cross-references: GB:AE001754; GB:AE000512; NID:q4981417; PIDN:AAD3  
A; Experimental source: strain MSB8  
C; Genetics:  
A; Gene: TM0893  
C; Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	13.2%	Score 162;	DB 2;	Length 246;
Best Local Similarity	33.3%;	Pred. No. 4.7e-06;		
Matches	48;	Conservative 23;	Mismatches 67;	Indels 6; Gaps 4;

  

Qy	93	LGADRLAAVGAHSLYPNTELLVIDAGTATYERVSAEGYVLGNISPGHLRFAKHLF	152
		: : : : :           : :           : : : : :	
Db	102	VGADRVANVAFVKEYGKNG - I I D M G T A T T V D L V - V N G S Y E G A I L P G F M V H S L P R G	159
Qy	153	TGRPLDTPSGISPKIAEYGSSTEEATTGVIHGLAGEIDRY IDLHAKEGSAVITGG	212
		: : : : :           : : : : : : : : : :	
Db	160	TAKLPLVE--VXPADPVVCKDTEENIRLGVNVSVALEGIIGRIKEYVGDLPVLTGG	216
Qy	213	DANYLARITRSGILIHDPVLVLGL	236
		: : : :     :	
Db	217	QSKIVKDMIKHEI - FDEDLTIKVY	239
		: : : :     :	

RESULT 14

BirA protein/bvg accessory factor NMB2075 [Imported] - Neisseria meningitidis (strain  
C/Species: Neisseria meningitidis  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C/Accession: B81009  
R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty,  
xi, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.  
A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A/Reference number: A81000; MUID:20175755; PMID:10710307  
A/Accession: B81009  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-592 <TET>  
A/Cross-references: GB:AEO002557; GB:AEO002098; NID:g7227332; PIDN:AAF42394.1; PID:g7277  
A/Experimental source: serogroup B, strain MC58  
C/Genetics:  
A/Gene: NMB2075

Query Match	13.1%	Score	161;	DB 2;	Length	592;			
Best Local Similarity	25.0%;	Pred. No.	1.8e-05;						
Matches	65;	Conservative	43;	Mismatches	102;	Indels	50;	Gaps	9;
Qy	5	LIYDOGNSACKVAFVRNNSIESIFLPKGAGQALSHLVAPHRFDKATYSSVGLPDEEAEA	64						
		:              : :   :							
Db	341	LLDDGNSRLKAWVENGTFATVGSAPYR---DLSPGAEWA-----EKADG	384						
		:              : :   :							
Qy	65	IVR--SCAAA-----SLMMGTETFPVPLRLQYDR--RTLGDRLAAVVGAAH	105						
		:              : :   :							
Db	385	NVIRVCGVGEFFKAQVQQLARKIEWLSPSSQAQALGIRNHYRPEHSGDWRFNALGSR	444						
		:              : :   :							
Qy	105	SLYPNTPELLVIDAGTATYVERVSAEGYIYLGAGNISPGIHLRFKALHLETFGRPLIDP	165						
		:              : :   :							
Db	445	RFSERNA-CVVVSCGTATVDALTDGHHYLGSTIMPGFHLMKESLAVRTANL-----NRHA	498						
		:              : :   :							
Qy	165	PKIAEYGSSTPEEITAGVIRHGLAGEIDRYIDDLHAKEGRSA---VILTGGDANYLARIIR	222						
		:              : :   :							
Db	499	GKRYPPFTTTGNVAVASGMDMAVCGSVNMHGRLEKKTGAGKPVVDVIITGGCAAKVAEALP	558						
		:              : :   :							
Qy	223	SGIL-----IHPDLVLLGL	236						
Db	559	PAFLAENTVRVADNLVIYGL	578						

Query Match	12.8%;	Score 157.5;	DB 2;	Length 274;
Best Local Similarity	31.5%;	Pred. No. 1.3e-05;		
Matches 58; Conservative 29; Mismatches 66; Indels 31; Gaps 9;				
QY	77	GTTPVPLRQYDRRTLGADRLAAVVGCAHSIYPNTLLVTDAGTAITVERVSARGIYVGG	136	
Db	93	GVRTGIPL-LVDNPKVEGADRIYVNCILAARFKF-QQAAIVDFGSSICVDVVSARKEFLGG	150	
QY	137	NISPGHLH-----RPKALHLFTGRPLIDPSPGISPKIAFYGSTBEATAGVHIGLAG	189	
Db	151	AIAPGVQVSDAAAAARSAAALR---RVELARPRSV-----VGKNTVECMQAGVVFVGFAG	200	
QY	190	EID-----RYIDDLHAKEG---RSAVITLTGSDANYLARIIRSGTILIHDP--LVLLGLNRI	239	
Db	201	LVDPGLVCRMQRDVEEFSGDGLGNRVAVVATGHTAPLLLPETHL---VDHYDRHLTTHGLRVL	258	
QY	240	LEYN 243		
Db	259	FERN 262		

Search completed: June 24, 2003, 22:03:22  
Job time : 9.8997 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.64387 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-15

Perfect score: 1392

Sequence: 1 MIIIDSGNSRLKVGWFFDP.....LDSPLVDGLAALAAQAQAPTA 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	267	1	BAF_BORPE
2	267	19.2	54	1	BAF_BORPE
3	125	9.0	233	1	YACE_BACSU
4	107.5	7.7	338	1	YMO5_MYCTU
5	106.5	7.7	423	1	VG75_HSV11
6	102.5	7.4	442	1	ARO_A_BORPE
7	101	7.3	1110	1	CYGD_BOVIN
8	99.5	7.1	700	1	PURL_HA1N1
9	98.5	7.1	442	1	ARO_A_BORPE
10	98.5	7.1	676	1	ICPO_HSVBK
11	94.5	6.8	449	1	SNDH_ACELI
12	92.5	6.6	440	1	GAT4_MOUSE
13	92	6.6	267	1	THIG_DEIRA
14	91	6.5	408	1	BPHG_BURCE
15	90.5	6.5	351	1	KLF2_RAT
16	90	6.5	759	1	YEHM_ECOLI
17	89.5	6.4	442	1	GAT4_HUMAN
18	89.5	6.4	507	1	VL2_HPV1A
19	89.5	6.4	559	1	MERA_ALCSP
20	89.5	6.4	747	1	ELS_BOVIN
21	89	6.4	1140	1	BCSC_ECOLI
22	88.5	6.4	361	1	COBT_MYCTU
23	88.5	6.4	1083	1	T2D3_HUMAN
24	88.5	6.4	3530	1	MY15_HUMAN
25	88	6.3	462	1	MURC_STECO
26	88	6.3	530	1	MORE_MYCLE
27	87.5	6.3	354	1	ODPA_ZYMO
28	87.5	6.3	459	1	COBG_PSEDE
29	87.5	6.3	1108	1	COGE_RAT
30	87.5	6.3	1625	1	CTPL_MYCTU
31	87.5	6.3	2205	1	POLN_RUBVT
32	87	6.2	465	1	EF3_HUMAN
33	87	6.2	631	1	Y282_MYCTU

RESULT 1  
BAF\_BORPE  
ID BAF\_BORPE STANDARD; PRT; 267 AA.  
AC Q45338; Q45373;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Bvg accessory factor.  
GN BAF.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID:520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BP504;  
RX MEDLINE=95325323; PubMed=7601846;  
RA Deshazer D., Wood G.E., Friedman R.L.;  
RT "Identification of a Bordetella pertussis regulatory factor required  
RT for transcription of the pertussis toxin operon in Escherichia  
RT coli.";  
RL J. Bacteriol. 177:3801-3807(1995).  
RN [2]  
RP SEQUENCE OF 1-38 FROM N.A.  
RC STRAIN-BP504;  
RA Wood G.E., Friedman R.L.;  
RT "Identification of a bira homolog in Bordetella pertussis.";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 239-267 FROM N.A.  
RC STRAIN-BP536;  
RX MEDLINE=96419162; PubMed=8821935;  
RA Allen A.G., Maskell D.J.;  
RT "The identification, cloning and mutagenesis of a genetic locus  
RT required for lipopolysaccharide biosynthesis in Bordetella  
RT pertussis.";  
RL Mol. Microbiol. 19:37-52(1996).  
CC -1. FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
CC A BVG-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
CC RNA POLYMERASE.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U12020; AAA75361.1; -  
DR EMBL; AF016461; AAC68834.1; -  
DR EMBL; X50711; CAA62242.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Transcription regulation; Activator.  
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

34 87 6.2 655 1 PRIA\_MYCTU P71670 mycobacteri  
35 86.5 6.2 226 1 BIOD\_MYCTU O06620 mycobacteri  
36 86.5 6.2 750 1 ELS\_CHICK P07916 gallus gall  
37 86 6.2 434 1 UL43\_HSV11 P10227 herpes simp  
38 86 6.2 516 1 AROA\_BRANA P17688 brassica na  
39 85.5 6.1 383 1 MRP\_MYCLE P53382 mycobacteri  
40 85.5 6.1 409 1 AMBL\_BACST P37113 bacillus st  
41 85.5 6.1 499 1 FEAB\_ECOLI P80668 escherichia  
42 85.5 6.1 1044 1 RSGI\_BOVIN P09851 bos taurus  
43 85 6.1 375 1 ACT1\_TETTH P10992 tetrahymena  
44 85 6.1 376 1 ACT2\_TETPY P0993 tetrahymena  
45 85 6.1 724 1 MALQ\_MYCTU O53932 mycobacteri

```

Query Match          100.0%; Score 1392; DB 1; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.6e-98;
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIIIDSGNSRLKVGWFPDPAQAAREPAPVAFNDLDDALGRWLATLPRRPQARLGVNV 60
DB 1 MIIIDSGNSRLKVGWFPDPAQAAREPAPVAFNDLDDALGRWLATLPRRPQARLGVNV 60
QY 61 AGLARGEIAATLRAGGCDIRWLRQAOLPLAMGLRNGYRNPDLQADRWACWGLARQPSV 120
DB 61 AGLARGEIAATLRAGGCDIRWLRQAOLPLAMGLRNGYRNPDLQADRWACWGLARQPSV 120
QY 121 HPPLIVASFGTATLDTTGPONVFPGLILPCPAMRGALAYGTAHLPLADGLVADYDIP 180
DB 121 HPPLIVASFGTATLDTTGPONVFPGLILPCPAMRGALAYGTAHLPLADGLVADYDIP 180
QY 181 THQATASGIAAQAAGAIYRQWLAGRQRYGOAPEIIVAGGWPVEVQAEERLLAVTGAAPG 240
DB 181 THQATASGIAAQAAGAIYRQWLAGRQRYGOAPEIIVAGGWPVEVQAEERLLAVTGAAPG 240
QY 241 ATPQTYLDSFVLGSLAALAAQAAGPTA 267
DB 241 ATPQTYLDSFVLGSLAALAAQAAGPTA 267

RESULT 2
BAF_BORBR
ID BAF_BORBR STANDARD; PRT; 54 AA.
AC 088005;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bvg accessory factor (Fragment).
GN BAF OR BLPSI.36C.
OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OC NCBI_TaxID=518;
FN [1]
RP STRAIN=CN7635E;
RC SEQUENCE FROM N.A.
RA Stevens K., Churche C.M., Badcock K.L., Parkhill J., Preston A.,
RA Maskell D.J., Barrell B.G.;
RL Submitted (AUG1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS ADDITIONAL FACTOR TO BVGAS TO ACTIVATE
CC TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ007747; CAA07675.1;
CC Transcription regulation; Activator.
FT NON_TER 1
SQ SEQUENCE 54 AA; 5447 MW; E0C44CC5107A8365 CRC64;

Query Match          19.2%; Score 267; DB 1; Length 54;
Best Local Similarity 96.3%; Pred. No. 3.6e-14;
Matches 52; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 214 IYVAGGWPVEVQAEERLLAVTGAAGTATPQTYLDSFVLGSLAALAAQAAGPTA 267
DB 1 IYVAGGWPVEVQAEERLLAVTGAAGTATPQTYLDSFVLGSLAALAAQAAGPTA 54

RESULT 3
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.

```

```

AC DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin."
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS Bvg ACCESSORY FACTOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D26185; BAA05305.1;
CC EMBL; Z99104; CAB11846.1;
CC Subtilist; BG10133; YACB.
CC InterPro; IPR004619; Baf.
CC Pfam; PF03309; Bvg_acc_factor; 1.
CC TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match          9.0%; Score 125; DB 1; Length 233;
Best Local Similarity 30.8%; Pred. No. 0.0085;

```



```

SQ  SEQUENCE  358 AA;  35592 MW;  AC7EDC5BF8E41544 CRC64;
      7.7%;  Score 107.5;  DB 1;  Length 358;
Query Match      Best Local Similarity 26.5%;  Pred. NO. 0.28;
Matches 75;  Conservative 30;  Mismatches 99;  Indels 79;  Gaps 15;

Oy  9 NSRLKGVG--FDPPAQAAREPAPVAFDNLDDLGRWLATLPRRPQORALGVNVAAGLARGE 67
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  59 NTVVNAAWFDPGSAFYLECA-----QACGLGLGGPPPTETALAHSKGV--GQ 117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy  68 AYAATFLRAGGCDIRWLRAQPLAMGLRNGYNPD-----QLG----ADRWACWVGVLAR 116
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  118 LTAALRLAG-----AARIVGVG--GGSACTGGKGMIAFLGIGDAARFOIADYEVTA 168
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Qy	117	QPSVHPPLLVASFSGTATTTLDTICPDNVFPGGLILPGPAMMKGALAYGTAHLPGLADGLVAD	176
Dd	159	SDVEYPT-LLGPWGTA-----RVF-----APQKA--DMATVAVLEGR <sub>LAA</sub>	205
Qy	177	YPIDTHQAISGIA-----AAQGAIVRQWLA--GRORYGQA-----PEIYV	216
Dd	206	WAIELDAAGRGVSAPGGAAGAIGAGLAVGGRYQSRAAIIAEHTHFADDLADAELIV	265
Qy	217	AGGWPEVRQEAERLLAVTGAFAFGTPQPQTLDSPVLGD <sub>LAA</sub> L	259
Dd	266	TGEG----RFDEQSLHGKVGVAIAAARPLAIPVIVLAGOVSL	304
 RESULT 5 VG75_HSV11            STANDARD;       PRT;     423 AA.			
AC	Q00128;		
DT	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	01-DEC-1992 (Rel. 24, Last annotation update)		
DE	Hypothetical gene 75 protein.		
GN	75		
OS	Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Ictalurid Herpes-like viruses.		
OX	NCBI_TaxID=10401;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Auburn 1;		
RX	MEDLINE=92087490; PubMed=1727613;		
RA	Davison A.J.;		
RT	"Channel catfish virus: a new type of herpesvirus.";		
RL	Virology 186:9-14(1992).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M75136; AAA88177.1; -		
DR	PIR; A36794; A36794.		
KW	Hypothetical protein.		
SQ	SEQUENCE 423 AA; 46531 MW; 07AC0FD9CEE0F7 CRC64;		
 Query Match                      7.7%; Score 106.5; DB 1; Length 423; Best Local Similarity    23.4%; Pred.No.0.39; Matches    50; Conservative    24; Mismatches    65; Indels    75; Gaps    11;			
Qy	56	LGNNVAGLARGEATAATLRAGGCIDRWLRAQLAMGLNGY-----	96
Dd	151	MGPVPQGMPT-ETISITINPFGGWSWLVLKPAGTSRKRPFFANDLTRYLIRTS <sub>G</sub> CFSTAGI	208
Qy	97	-----RNPDL-GADRWCAMGVGLAQPSVHPPLLVASGCTATLTDTIGPDNVPFGUIL	150
Dd	209	DRFETRSPILIYGGRW-----VLCVPHLHPTLTVAKESECALD-----PGSIW	255

```
QY 151 --PGAMRGALAYCTAHLPLADGLVADYPIDTHQAIASGIAA-----AQAGAIYR 199
Db 256 IWPGEA-----LPVVRHVLETVLVDVRDLVGTLANFEDIFVHRVDRRVTH 303
QY 200 QWL-----AGRQY-----GQAPETIYVA--GGGW 221
Db 304 RWEIPKPNAPRKLSFTVTGHRPEVLAAMCGGGW 337

RESULT 6
ARO_A_BORPE
ID ARO_A_BORPE STANDARD; PRT; 442 AA.
AC P12421.
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OC NCBI_TaxID=520;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88227818; PubMed=2897356;
RA Maskell D.J., Morrissey P., Dougan G.;
RT "Cloning and nucleotide sequence of the aroA gene of Bordetella
RT pertussis.";
RL J. Bacteriol. 170:2467-2471(1988).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate -
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20023; AAA22968.1;
DR PIR; A32007; XUBRV5.
DR InterPro; IPR001986; EPSP_syntase.
DR Pfam; PF00275; EPSP_syntase; 1.
DR ProDom; PD001867; EPSP_syntase; 1.
DR PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferrase.
SQ SEQUENCE 442 AA; 46588 MW; D81088B36D967955 CRC64;

Query Match 7.4%; Score 102.5; DB 1; Length 442;
Best Local Similarity 26.2%; Pred No. 0.82;
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQALGVNVAGLAR-----GEAIAATLRAGGCDIR----- 81
Db 115 LSGVPRMHERPIGLVDALRQFAGIEVLGQAGVPPPLRIGGSIKRVDPVRVEGVSQF 174
QY 82 ---NLRAQPLANGLRNGYRNPQDLGADRWACWGVLAQPSVHPPL-LVASFQATTLLD- 136
Db 175 LTALLMAAPV-LARRS-----GQDITIEVWGLISKPIETITLNLMAFVGVSVRDG 225
QY 137 ----TIGPDNVF--PGGLILPGPAMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183
Db 226 WRATFIADAVYRPGRWAIIGDA-----STASYFLALGATGGGPRVTGVEDSIQ 277
QY 184 A-IASGIAAQAQAIVRQWLQAGRQYGOAPEIYVAGGQWPEVR-----QEAERLLA----- 233
```

```
Db 278 GDVFAATLAAMGADVRY-----GPGWIEGRVVRVAEGRLKAFDAF 320
QY 234 --VTGAAGFATQPTLYLDS 251
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 7
CYGD_BOVIN
ID CYGD_BOVIN STANDARD; PRT; 1110 AA.
AC P55203; 002809;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retinal guanylyl cyclase 1 precursor (EC 4.6.1.2) (Guanylate cyclase
DE 2D, retinal) (RETGC-1) (Rod outer segment membrane guanylate cyclase)
DE (ROS-GC) (Guanylate cyclase E) (GC-E).
GN GUCY2D OR GUC2D.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94379976; PubMed=7916565;
RA Goraczniak R.M., Duda T., Sitaramayya A., Sharma R.K.;
RT "Structural and functional characterization of the rod outer segment
RT membrane guanylate cyclase.";
RL Biochem. J. 302:455-461(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Duda T., Venkataraman V., Sharma R.K.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97398155; PubMed=9256080;
RA Johnston J.P., Farhangfar F., Aparicio J.G., Nam S.H., Applebury M.L.;
RT "The bovine guanylate cyclase GC-E gene and 5' flanking region.";
RL Gene 193:219-227(1997).
RN [4]
RP SEQUENCE OF 57-71; 163-178; 702-713 AND 749-763.
RX TISSUE=Retina;
RA Margulis A., Goraczniak R.M., Duda T., Sharma R.K., Sitaramayya A.;
RT "Structural and biochemical identity of retinal rod outer segment
RT membrane guanylate cyclase.";
RL Biochem. Biophys. Res. Commun. 194:855-861(1993).
CC -1- FUNCTION: PROBABLY PLAYS A SPECIFIC FUNCTIONAL ROLE IN THE RODS
CC AND/OR CONES OF PHOTORECEPTORS. IT MAY BE THE ENZYME INVOLVED IN
CC THE RESYNTHESIS OF GMP REQUIRED FOR RECOVERY OF THE DARK STAGE
CC AFTER PHOTOTRANSDUCTION.
CC -1- CATALYTIC ACTIVITY: GTP = 3',5'-cyclic GMP + diphosphate.
CC -1- ENZYME REGULATION: INHIBITED BY ATP-MG OR ATP-MN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE RETINA.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L37089; AAA50790.1;
CC EMBL; AF027203; AAB86385.1;
DR
```



```

GN AROA.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=240/2;
RA "Arthur J.D., Walker M.J.;
RT "Bordetella bronchiseptica aroA";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate -
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -|- SUBUNIT: MONOMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -|- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF182427; AAF01290.1; -
CC InterPro; IPR001986; EPSP_synthase.
CC Pfam; PF00275; EPSP_synthase; 1.
CC ProDom; PD001867; EPSP_synthase; 1.
CC PROSITE; PS00104; EPSP-SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP-SYNTHASE_2; 1.
CC Aromatic amino acid biosynthesis; Transferase.
CC SEQUENCE 442 AA; 46629 MW; 758313D594D2F00B CRC64;
Query Match 7.1%; Score 98.5; DB 1; Length 442;
Best Local Similarity 25.8%; Pred. No. 1.6;
Matches 67; Conservative 22; Mismatches 84; Indels 87; Gaps 13;
QY 45 LATIPRRPQALGVNVNGLAR-----GEAATLRRAGGDIR----- 81
DB 115 LSGVPRMERHIGDLDVLRFGAGIEYLGQAGYPLRIGGSIRVDGPRVGVSSQF 174
QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRWACWGVGLARQPSVHPPL-LVASFGTATLTD- 136
DB 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYEITLNLMAFGVSVRRDG 225
QY 137 ---TIGPDNVF--PGGLILPGPMMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183
J 226 WRAFTIARDAAYRGPRMAIEGDA-----STASYFLALGAIGGPPVVTGVGEDSIQ 277
QY 184 A-IASGIAAQAAGAIVRQWLAGRQRYGQAPETIYVAGGWPVR-----QEAELLA----- 233
DB 278 GDVAFAATLAAMGADVY-----GPGWIETRGVRVAEGGRLKAFDAF 320
QY 234 --VTGAFAFGATPQPTYLDSP 251
DB 321 NLIPDAAMTAATLALYADGP 340
RESULT 10
ICPO_HSVBK
ID ICPO_HSVBK STANDARD; PRT; 676 AA.
AC P29836;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trans-acting transcriptional protein ICPO (P135 protein) (IER
DE 2.9/ER2.6).
GN BICPO.
OS Bovine herpesvirus type 1 (strain K22).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

```

```

OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31519;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219360; PubMed=1313901;
RA Wirth U.V., Fraefel C., Vogt B., Vicek C., Paces V., Schwytzer M.;
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
RT are 3' coterminal and encode a putative zinc finger transactivator
RT protein.";
RL J. Virol. 66:2763-2772(1992).
CC -|- PTM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
CC ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
CC CASIN KINASE II.
CC -|- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -|- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84464; AAA46061.1; -
CC PIR; A38209; EDBE22.
CC HSSP; P28990; LCHC.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00097; zf-C3HC4; 1.
CC SMART; SM00184; RING; 1.
CC PROSITE; PS00518; ZF_RING_1; 1.
CC PROSITE; PS00889; ZF_RING_2; 1.
CC Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
CC DNA-binding; Early protein; Repressor; Phosphorylation.
CC ZN_FING 13 52 RING-TYPE.
CC FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC).
CC SEQUENCE 676 AA; 67701 MW; 9BB0683C9BFC65D CRC64;
Query Match 7.1%; Score 98.5; DB 1; Length 676;
Best Local Similarity 24.6%; Pred. No. 2.5;
Matches 67; Conservative 17; Mismatches 85; Indels 103; Gaps 13;
QY 21 APQAREPAPVAFDNLDDLALGRWLATLPRR-----PQALGVNVNGLARGEIAATLRA 75
DB 476 APAPAPAPIS-TMIDL-----TANAPAPADPAPAAALGPALAGA----- 516
QY 76 GGCIDRWLRQAQPLAMGLRNGYRNPDLGADRWACWGVGLARQPSV-----HPPLLVSF 129
DB 517 -----QIGTPAAAATAATAAAPSVAAGSAPSAVTAAT 551
QY 130 GTATTLDITGPDNVFPFGLILPGPMMRGALAYGTAHLPLADGLVADYPIDTHQATASGI 189
DB 552 GTAAALSTRAP-----TPSPA-----GRAPADPRRAGAP-----ALA-GA 586
QY 190 AAAQAAGAIVRQWLAGRQRYGQAPETIYVAGG---WPEVROEAEERLLAVTGAAP----- 239
DB 587 ARAEAG---RNGNPGRER---RPASAMRGDLDPGPESSAQKRRRTTEMEVAANVRESLLG 640
QY 240 -----GATPQPTYLDSPVLDGLAALAQA 264
DB 641 TPRSSAALAPQPGRGQPSLAGLLGRCSGGS 672
RESULT 11
SNDH_ACELI
ID SNDH_ACELI STANDARD; PRT; 449 AA.
AC Q44091;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE L-sorbose dehydrogenase (EC 1.1.1.-) (SNDH)
OS Acetobacter liquefaciens (Gluconacetobacter liquefaciens).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;

```



Db 42 PRVPSVGLSYL-QGGSSAAAGTSGSS-----GAGPSGAGPGTQGGSPGSHQAGAE- 95

QY 107 WACWGVLARQPSVHPPLLVASGATATLDTIGPDNVFPGG-----LILPGPAMMRGALAY 162

Db 96 -----GAAATPPVPSPRFSFGCTTGLSAAAAAARAAAY 132

QY 163 GTAHLPLADGLVADYDTHQATASGIAAQAQAIAVROWLAGRQRYGQ-----AP-E 213

Db 133 G-----SGGAAGAG-----LAGREQRGPGFAGSYSPYP 163

QY 214 IYVA--GGGWPEVRQBAERLLAVTGAAGATPOPTVLDSPVLDGLAALAAQ 263

Db 164 AYMAVAGSW-----AAAAASAGP--FDSFVLSLPGRANPG 199

RESULT 13

THIG.DEIRA

ID THIG.DEIRA STANDARD; PRT; 267 AA.

AC Q9RYV1

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Thiazole biosynthesis protein thig.

GN THIG OR DRA0172

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI\_TaxID=1299;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=R1

RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.

RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1."

RL Science 286:1571-1577(1999).

CC -!- FUNCTION: Required for the synthesis of the thiazole moiety of

CC thiamine (By similarity).

CC -!- COFACTOR: FMN (Potential).

CC -!- PATHWAY: Thiamine biosynthesis.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: BELONGS TO THE THIG FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE001862; AAF12198.1; ALT\_INIT.

DR TIGR; DRA0172; -

DR InterPro; IPR003009; FMN enzyme.

KW Thiamine biosynthesis; Flavoprotein; FMN; Complete proteome.

SQ SEQUENCE 267 AA; 27590 MW; A4ED02AD3D7DAA98 CRC64;

Query Match

Best Local Similarity 6.6%; Score 92; DB 1; Length 267;

Matches 65; Conservative 21; Mismatches 70; Indels 68; Gaps 15;

QY 34 DNLDDALGRWLATLP-----RRPQALGVNVAGLARGEATAATLRAGGCDIRWLRAQPL 88

Db 62 DALDLDR-----FQLLPNTAGCTAEEA--VRVAKLAR-----AAT-----GVSWLKLEVI 105

QY 89 AMGLNGRYRNPDLQADRWACWV-----GVLARQ-----PSVHPPLLVAS 128

Db 106 P---DPKYLIPDFIGTIRAAEILVGEFTVLPYVPGDGLARALEAVCCATVMP---LASP 160

QY 129 FGTA-----TTLDTIGPDNVFPGGILPQPMRGALAYGTAHLPLADGLVADY- 177

Db 161 IGTGKGLRTPELINTVIDGAGVPIVDAGLGVPSDAQ--ALELG-ADAVLVNTAIEAR 217

QY 178 -PIDTHQATASGIAAQAQAIAVROWLAGR--ORYGQAPEIYVAG 218

Db 218 DPVAMARAFALGVENGRLG-----YLAGRMQQRDSASPSPAAG 256

RESULT 14

BPHG.BURCE

ID BPHG.BURCE STANDARD; PRT; 408 AA.

AC P37337;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Biphenyl dioxygenase system ferredoxin--NAD(+) reductase component

DE (EC 1.18.1.3).

GN BPHG.

OS Burkholderia cepacia (Pseudomonas cepacia).

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Burkholderia.

OX NCBI\_TaxID=292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LB400;

RX MEDLINE=92234948; PubMed=1569021;

RA Erickson B.D., Mondello F.J.;

RT "Nucleotide sequencing and transcriptional mapping of the genes

RT encoding biphenyl dioxygenase, a multicomponent

RT polychlorinated-biphenyl-degrading enzyme in Pseudomonas strain

RT LB400."

RL J. Bacteriol. 174:2903-2912(1992).

RN [2]

RP REVISIONS.

RA Erickson B.D., Mondello F.J.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BIPHENYL

CC DIOXYGENASE, TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO

CC NADH.

CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NAD(+) -> oxidized

CC ferredoxin + NADH.

CC -!- COFACTOR: FAD.

CC -!- PATHWAY: Biphenyl and polychlorinated biphenyl degradation.

CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE

CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHE), A

CC FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).

CC -!- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE

CC FERREDOXIN REDUCTASE COMPONENTS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M86348; AAB63429.1; -

DR F1858; F41858.

DR InterPro; IPR001327; FAD\_pyr\_redux.

DR InterPro; IPR000205; NAD\_binding.

DR Pfam; PF00070; pyr\_redux; 1.

DR PRINTS; PR00368; FADPNR.

DR ProDom; PD000139; FAD\_pyr\_redux; 1.

KW Aromatic hydrocarbons catabolism; Flavoprotein; FAD; NAD;

KW Oxidoreductase.

FT NP\_BIND 4 35 FAD (ADP PART) (POTENTIAL).

FT NP\_BIND 145 173 NAD (ADP PART) (POTENTIAL).

SQ SEQUENCE 408 AA; 42953 MW; 8A52BB01688667A9 CRC64;

```

RESULT 15
KLF2_RAT
ID KLF2_RAT STANDARD; PRT; 351 AA.
AC Q9ET58;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE kruppel-like factor 2 (Lung kruppel-like factor).
KLF2 OR LKLF.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
STRAIN=BN/SSNHSD;
RA Haag F., Bartels K., Rothenburg S., Stahmer I., Thiele H.-G.,
RT Koch-Nolte F.;
RT "The gene for the transcription factor LKLF is developmentally
RT expressed in rat T cells and is not defective in lymphopenic
RT diabetes-prone BB rats.";
RT Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
"-L" C
"-I" FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
CC AND ACTIVATES TRANSCRIPTION (BY SIMILARITY).
CC "-I" SUBCELLULAR LOCATION: Nuclear (By similarity).
CC "-I" SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF181251; AAC02141.1; .
CC HSP; P08047; ISP2.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
CC DNA-binding; Nuclear protein; Repeat.
CC DOMAIN 60 70 POLY-PRO.
CC FT DOMAIN 165 168 POLY-PRO.

```

**THIS PAGE BLANK (USPTO)**



Result No.	Score	Query			DB	ID	Description
		Match	Length	DB			
1	311	22.3	295	16	Q8Y2M4	Q8Y2m4	<i>ralstonia s</i>
2	256.5	18.4	592	16	Q9JW17	Q9Jw17	<i>neisseria m</i>
3	254.5	18.3	592	16	Q9JXF1	Q9Jxf1	<i>neisseria m</i>
4	194.5	14.0	242	16	Q9PC14	Q9pc14	<i>xylella fas</i>
5	178.5	12.8	248	16	Q9HWG1	Q9hwcl	<i>pseudomonas</i>
6	161	11.6	259	16	Q8YAC5	Q8yac5	<i>listeria mo</i>
7	160.5	11.5	273	16	Q9YB84	Q97eb4	<i>clostridium</i>
8	156	11.2	255	16	Q8R7M2	Q8r7m2	<i>thermoanaer</i>
9	155	11.1	259	16	Q2PF54	Q92f54	<i>listeria in</i>
10	152	10.9	246	16	Q9WXY5	Q9wzy5	<i>thermotoga</i>
11	138	9.9	212	2	Q32514	Q32514	<i>desulfotibr</i>
12	134.5	9.7	261	16	Q9A6Z1	Q9a6z1	<i>caulobacter</i>
13	134	9.6	265	16	Q9XN6	Q9x8n6	<i>streptomyce</i>
14	131.5	9.4	254	16	Q9KGH5	Q9kg85	<i>bacillus ha</i>
15	128.5	9.2	259	16	Q8XHL5	Q8xhl5	<i>clostridium</i>
16	124.5	8.9	258	2	Q9F985	Q9f985	<i>bacillus st</i>

```

QY 46 ATLPRRPGALGVNAGLARGAIAATLRA--GGCDIRWLRAQPLAMGLRNGYRNPDLG 103
DB 72 AP-----PDWISV-VAGPALRDALCARIVFDGARLRIVASEAAAGLRNGYRDPAGL 126
QY 104 ADRWACWGLVARQPSVHPPLIVASFGTATTITDTPONVFPGLILPGPMARGALAYG 163
DB 127 TDRWVGAVG--ARHAWPDTALLVTAGTATLIDIVAPDGRFAGGLILPGLTLMRALS 184
QY 164 TAHLPLAD-GLV-----ADYPI---DTHQIASGIAAQAQAIAIVRWLAGRQRYGQ 210
DB 185 TAQLPEIDIGVLAARDDAQADVPDSWADNTODATLALGCVTAQAGAIAGTQALQAQY 244
QY 211 APEIIVAGG---WPEVROEAERLLAVTGAFAFGTPOPTIYLDSPVLDGLAALAAQ 263
DB 245 PYRCVLGGARAALAPHLRMPFQ-----MHDNLVLGLQVLAHTG 284

RESULT 2
QJWI7 PRELIMINARY; PRT; 592 AA.
QJWI7
JT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Bira bifunctional protein (EC 6.3.4.15).
GN BIRA OR NMA0357.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moutle S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RL meningitidis 2491."
RL Nature 404:502-506(2000).
DR EMBL; AF062752; CAB83659.1; -.
DR HSSP; P06709; IBIA.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; Bira_ligase.
DR InterPro; IPR003142; BPLC.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPLC; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR TIGRFAMs; TIGR00121; bira_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
DR Ligase; Complete proteome.
SQ SEQUENCE 592 AA; 64468 MW; 7ED2E0CD3B31C630 CRC64;

Query Match 18.4%; Score 256.5; DB 16; Length 592;
Best Local Similarity 32.0%; Pred. No. 3.1e-10;
Matches 85; Conservative 43; Mismatches 113; Indels 25; Gaps 9;

QY 3 ILIDSGNSRLKVGWFPDPAQAREPAPVAFDNLDLALGRWLATLPRRPGALGVNAG 62
DB 341 LLLDGGNSRLKAWVE-NGTFATVGSAPYR----DLSPLGAWEAEKVDGNRIYGCACVG 395
QY 63 LARGEATAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACWGLVARQPSVHP 122
DB 396 EFKAQVQEQAR---KIEWLPSSAQALGIRNHYRHPHEGSDRWFNALG--SRFS-RN 449
QY 123 PLLVASFGTATLTDTIGPDNVFPGLILPGPMARGALAYGTAHLPLADGLVADYPIPTH 182
DB 450 ACVVVSGTAVTVDALTDGHLGTTMPGFHLMKESLAVRTANLRHAGKRYPPPTTGT 509
QY 182 PLLVASFGTATLTDTIGPDNVFPGLILPGPMARGALAYGTAHLPLADGLVADYPIPTH 182
DB 450 ACVVVSGTAVTVDALTDGHLGTTMPGFHLMKESLAVRTANLRHAGKRYPPPTTGT 509

```

```

QY 183 QAIASGIAAQAQAIAIVRWLAGRQR-----YQAPETIYVAGGWPVEVROEAERLLAVTCAA 238
DB 510 NAVASGMDAVCGSVN--MMHGRLEKTKTGAKPVDVIITGGG-----AAKVAEALPPA 560
QY 239 FGATPQPTIYLDSPVLDGLAAL-AAQG 263
DB 561 FLAENTVRVADNLVIHGLNLIAAEG 586

RESULT 3
QJXF1 PRELIMINARY; PRT; 592 AA.
QJXF1
AC QJXF1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Bira protein/Bvg accessory factor.
GN NMB2075.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2017555; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RT Science 287:1809-1815(2000).
RL EMBL; AF002557; AAF42394.1; -.
DR HSSP; P06709; IBIA.
DR TIGR; NMB2075; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; Bira_ligase.
DR InterPro; IPR003142; BPLC.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPLC; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00121; bira_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 592 AA; 64701 MW; 9E27421DA2B41DE3 CRC64;

Query Match 18.3%; Score 254.5; DB 16; Length 592;
Best Local Similarity 31.6%; Pred. No. 4.3e-10;
Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps 9;

QY 3 ILIDSGNSRLKVGWFPDPAQAREPAPVAFDNLDLALGRWLATLPRRPGALGVNAG 62
DB 341 LLLDGGNSRLKAWVE-NGTFATVGSAPYR----DLSPLGAWEAEKADGNRIYGCACVG 395
QY 63 LARGEATAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACWGLVARQPSVHP 122
DB 396 EFKAQVQEQAR---KIEWLPSSAQALGIRNHYRHPHEGSDRWFNALG--SRFS-RN 449
QY 123 PLLVASFGTATLTDTIGPDNVFPGLILPGPMARGALAYGTAHLPLADGLVADYPIPTH 182
DB 450 ACVVVSGTAVTVDALTDGHLGTTMPGFHLMKESLAVRTANLRHAGKRYPPPTTGT 509
QY 183 QAIASGIAAQAQAIAIVRWLAGRQR-----YQAPETIYVAGGWPVEVROEAERLLAVTCAA 238
DB 510 NAVASGMDAVCGSVN--MMHGRLEKTKTGAKPVDVIITGGG-----AAKVAEALPPA 560

```



```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Blocher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkut G.,
RA Madieno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
A Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
AT *Comparative genomics of Listeria species.*;
RL Science 294:849-852(2001).
DR EMBL: AL591974; CAD00748.1; -.
DR MEROPS: M41.009; -.
DR ListLst; LMO00221; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein. Complete proteome.
SQ SEQUENCE. 259 AA; 28187 MW; F10B049D80025B7 CRC64;

Query Match 11.6%; Score 161; DB 16; Length 259;
Best Local Similarity 26.3%; Pred. No. 0.00062;
Matches 78; Conservative 38; Mismatches 109; Indels 72; Gaps 16;

QY 1 MIIIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRW-LATLPRPQALGV- 58
DB 1 MIIIDVGNVTGTVGYEKQ-----KLLKHWRTTDRHRTSDELGMT 42
QY 59 -----NVAGLARGEIAATLRA--GGDIRWLRAQPLAM--GLRNGYR--- 97
DB 43 VLNFFSYANLTSDTQGIISVVPIHMETMC--VFNFNRLPLVGGIKTGLMLKVD 101
QY 98 NPDQLGADRWACWGVGLARQSPVHPPLLVASFGTATTLTIGPDNVFPGLILPGPAMMR 157
DB 102 NPREIGSDR---IVNAVAASEEYGPVIVDFGTATTCYIDSGVYQGAIAPIGIMIST 158
QY 158 GALAYGTAHLPLADGLVADYPTDTHQAI A-SGIAAQAQ---AIVRQW---LAGRQRYGQ 210
DB 159 EALYNRAAKLPVDD--IA-----ESSQIIIGKSTVSSMQAGIFGVGQCCEGIIAEMKKQSN 212
QY 211 APEIIVAGGWEVPEVQAEERLLAVTGAAPGAPQPTLYDS--PVLDGLAALAAQAGAPT 266
DB 213 ASPVVATGGL--ARMIETKSSAV-----DILDPFUTLKGLELLYRNKPT 256

RESULT 7
Q97EB4 PRELIMINARY; PRT; 273 AA.
AC Q97EB4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Predicted transcriptional regulator, homolog of Bvg accessory
DE factor.
GN Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Bregon G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007815; AAK81136.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBR; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64;

Query Match 11.5%; Score 160.5; DB 16; Length 273;
Best Local Similarity 24.7%; Pred. No. 0.00072;
Matches 64; Conservative 42; Mismatches 104; Indels 49; Gaps 9;

QY 1 MIIIDSGNSRLKVGWFDPAQAPAPVAFDNLDLALGRW-LATLPRPQALGVN 59
DB 12 VILVLDVGNITVILGIYN-----DTKLTAEWRSLTDVLRSADEYGIQ 53
QY 60 VAGLARGEIAATLRA-----GDIRWLRAQPLAM--GLRNG---YRN 98
DB 54 VMNLFODKLDPTLVEGVIISSVVPNIMYSLHMKRYKINPLVVGPIKGINIKYDN 113
QY 99 PDQLGADRWACWGVGLARQSPVHPPLLVASFGTATTLTIGPDNVFPGLILPGPAMMRG 158
DB 114 PKEVGADR---IVNAVAHEIYKRSLLIIDGTATTCFAVRENGDYLGAICPGIKVSSE 170
QY 159 ALAYGTAHLPLADGLVADYPT--DTHQAIASGIAAQAQ---TVRWLAGRQRYGQAP 213
DB 171 ALFEKAALPFRVELIRPAPALCAKNTISSIOSGIVYGVQRYIVERNKEELQEGEKEP 230
QY 214 IYVAGGWPPE-VRQEAERL 231
DB 231 LVVATGGLAKLISEAKNV 249

RESULT 8
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013180; AAM2520.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 11.2%; Score 156; DB 16; Length 255;
Best Local Similarity 25.9%; Pred. No. 0.0014;

```





```
DR EMBL; AL049628; CAB40880.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; Baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match          9.6%; Score 134; DB 16; Length 265;
Best Local Similarity 25.8%; Pred. No. 0.051;
Matches 59; Conservative 31; Mismatches 83; Indels 56; Gaps 8;

QY 1 MILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRW-LATLPRRQALGVN 59
DB 1 MLITDVGNTHTVLGLDFGE-----DIVEHWRISTDSRRTADELAVL 42
QY 60 VAGL-----ARGEIAATL-----RAGGCDIRWLRAQPLAMGLRN 94
DB 43 LOGLMGMHPLLGDELGDGIDGAIACATVPSVLHELREVTTRYCDVPAVLVEP---GVKT 99
Y 95 GY----RNPDOLGADRWACMGVGLARQPSVHPPLLVASFATLTLDIGDNPVPGGLIL 150
DB 100 GVPILTDHPKEVGADR---IINVAVALYGGPAIVVDFTATFDAVSARGEYIGGVIA 156
QY 151 PGPMARGALAYTAHLPLADGLIVADYPI--DTHQAIASGIAAAQAAGAI 197
DB 157 PGTEISVEALGVKGALRKIEVARPSRVIGKNTVEAMQSGIVYGFAGQV 205

RESULT 14
Q9KGH5 PRELIMINARY; PRT; 254 AA.
ID O9KGH5 PRELIMINARY; PRT; 254 AA.
AC O9KGH5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.J.C.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
TA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; Baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match          9.4%; Score 131.5; DB 16; Length 254;
Best Local Similarity 24.1%; Pred. No. 0.072;
Matches 63; Conservative 43; Mismatches 98; Indels 57; Gaps 12;

QY 1 MILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRW-LATLPRRQALGVN 59
DB 1 MILVIDGNTVTLGVYQ-----DETIVHWRILATSQKTEDEVAMT 42
QY 60 V-----AGL-----ARGEIAATLRAGGCDI-----RWLRAQPLAM--GLRNG-----YRN 98
DB 43 VRSLFDHAGLQFODIDGIVISSVPPPMFSLQCMCKYFHVFTMIIGPGIKTGLNIKYD 102
QY 99 PDOLGADRWACMGVGLARQPSVHPPLLVASFATLTLDIGDNPVPGGLILPGPMARG 158
DB 103 PKEVGADR---IYNVAVAIELYGPAIVVDFTATFATTCYLLINEKKQYAGGVIAFGIMISTE 159
```

```
QY 159 ALAYGTAHLPLAD-----GLVADYPIDTHQAIASCI---AAQAAGAIYVROWLAGRQRYGQ 210
DB 160 ALYHRASKLPRIETAKPKQVVTWTIDSMQ---SGIFYGVYQVDVVVKRKAQAE---S 213
QY 211 APEIYVAGGGWPEVRQEAERL 231
DB 214 EPKVIATGGLAKLIGTSETI 234
```

## RESULT 15

```
Q8XHL5 PRELIMINARY; PRT; 259 AA.
ID O8XHL5 PRELIMINARY; PRT; 259 AA.
AC O8XHL5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRfams; TIGR00671; Baf; 1.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;
```

```
Query Match          9.2%; Score 128.5; DB 16; Length 259;
Best Local Similarity 23.3%; Pred. No. 0.12;
Matches 61; Conservative 43; Mismatches 119; Indels 39; Gaps 11;

QY 1 MILIDSGNSRLKVGWFDPD---APQAAREPAPVAFDNLDLALGRWLATLPRRQALG 57
DB 1 MILIDVGNTHVLGIHDKNEYIASWRISTDSKTSDEYSIQVM-QLFNQAALNPEDVEG 59
QY 58 VNVAGLARG--EATAATLRAGGCDIRWLRAQPLAM--GLRNG-----YRNPDLGADRWAC 109
DB 60 IITSSVVPNTIMHSLNNMVRKFC-----KEPIVVGPGIKTGINTKYNPKEVGADR--- 110
QY 110 MGVGLARQPSVHPPLLVASFATLTLDIGDNPVPGGLILPGPMARGALAYGTAHLPL 169
DB 111 IYNVAFAFERKKPMIILDFGTATTFCAITKGDYLGNGICPGIOISADALFERAKLPR 170
QY 170 AD-----GLVADYPIDTHQAIASGIAAAQAAG---IVROWLAGRORYGOAPEIYVAGGW 221
DB 171 IELEKPKSVCKNTVTSMQA---GIIVGYICKVEIVYVKKMKKMDMLGEKPEFVLATGSL 227
QY 222 -----PEVRQEAERLLAVTG 236
DB 228 AKLYSETDVIDEYDRKLTLEG 249
```

Search completed: June 24, 2003, 21:59:34  
Job time : 23.7766 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.7999 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MILLIDSGNSRLKVGWFDPPD.....LDSPLVDGLAALAAOQAPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	7.5	1841	2	US-08-804-227C-6
2	105	7.5	5215	4	US-09-105-537-2
3	102.5	7.4	440	6	5310667-9
4	102.5	7.4	442	1	US-08-476-008-64
5	102.5	7.4	442	1	US-08-306-063-64
6	102.5	7.4	442	1	US-08-833-485-64
7	102.5	7.4	442	4	US-09-137-440-64
8	102.5	7.4	1996	2	US-08-804-227C-9
9	102.5	7.4	1996	2	US-08-804-198-3
10	100	7.2	4472	2	US-08-804-227C-2
11	95.5	6.9	442	6	5310667-6
12	93.5	6.7	3729	2	US-08-804-227C-4
13	91	6.5	820	2	US-08-374-483-2
14	91	6.5	820	2	US-08-374-483-5
15	91	6.5	1864	2	US-08-804-227C-3
16	90.5	6.5	453	2	US-08-851-088-8
17	90.5	6.5	453	4	US-08-851-089-2
18	90	6.5	1611	2	US-08-804-227C-5
19	90	6.5	1891	2	US-08-804-227C-12
20	90	6.5	1891	2	US-08-804-198-6
21	89.5	6.4	4545	2	US-08-804-227C-14
22	89.5	6.4	4550	2	US-08-804-227C-8
23	89.5	6.4	4550	2	US-08-804-198-2
24	89	6.4	4150	4	US-09-428-517-2
25	88	6.3	3170	4	US-09-036-987A-4
26	88	6.3	3170	4	US-09-370-700-4
27	87.5	6.3	2205	1	US-08-093-453B-2

```

28 87 6.2 444 4 US-09-000-062-5 Sequence 5, Appli
29 87 6.2 570 3 US-08-832-078-5 Sequence 5, Appli
30 86.5 6.2 822 2 US-08-939-002A-16 Sequence 16, Appli
31 86 6.2 423 4 US-09-243-374-6 Sequence 6, Appli
32 86 6.2 444 1 US-08-476-008-51 Sequence 51, Appli
33 86 6.2 444 1 US-08-306-063-64 Sequence 51, Appli
34 86 6.2 444 1 US-08-833-485-64 Sequence 51, Appli
35 86 6.2 444 4 US-09-137-440-64 Sequence 51, Appli
36 86 6.2 3816 4 US-09-428-517-3 Sequence 3, Appli
37 86 6.2 5087 4 US-09-144-085-1 Sequence 1, Appli
38 86 6.2 6095 4 US-09-144-085-2 Sequence 2, Appli
39 84.5 6.1 877 2 US-08-407-875-2 Sequence 2, Appli
40 84.5 6.1 877 3 US-09-126-280-2 Sequence 2, Appli
41 84.5 6.1 877 4 US-09-277-858-2 Sequence 2, Appli
42 84.5 6.1 1580 2 US-08-804-227C-11 Sequence 11, Appli
43 84.5 6.1 1580 2 US-08-804-198-5 Sequence 5, Appli
44 84.5 6.1 3567 2 US-07-642-734C-4 Sequence 4, Appli
45 84.5 6.1 3567 3 US-08-439-009A-4 Sequence 4, Appli

```

## ALIGNMENTS

```

RESULT 1
US-08-804-227C-6
; Sequence 6, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1841 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-804-227C-6

```

Query Match 7.5%; Score 105; DB 2; Length 1841;

Best Local Similarity 24.6%; Pred.No. 0.031;  
Matches 87; Conservative 24; Mismatches 105; Indels 138; Gaps 20;

```

Qy 36 LDLD-----ALGRWLATLP-----RRPQAL-----GVNV 60
      |||||  |||||  |||||  |||||
Db 684 LDLDACALVAARGLMORLPGGAMVSVRAGEDVRLLAGREDVAVCVAVNGPSVVI 743
      |||||  |||||  |||||  |||||
Qy 61 AGLARGEAIATLRAG-GCDIRWLRA-----QPLAMGLRNGYR----- 97

```

Db 744 SGAEEAEEAQAAGRRTRRLRVAAHAFHSPMDMLAGFVEAAGLRYPETLVVS 803  
QY 98 -----NPDOL-GADRWACMGVILARQP-----SVHPDLLVASFGTATTLTIGPD 141  
Db 804 TVTGRPARPGELTGDYWAQV-----REPVRFADAVRTAH-----RLGARTFLET-GPD 852  
QY 142 NVF-----PGGLILPGPAMRGALAYGTAHLPLADGLVAD 176  
Db 853 GVLGMAEECEDDDTVALLPAHKPGTAPHGPAAPG-ALRAAAAYGRGARVDWAGMHAD 911  
QY 177 YPIDTHOAIASGIAAQAAGAIWOMLA-CR-----ORYGOAPEIYVAGGG-----WP 222  
Db 912 GPEGPARVELPVHAFRHR---RYWLAPGAAATDDWYRIGWDRPLPATVGGARTAGRWL 968  
QY 223 EVROEAERLLAVTG-----AAGFATPQPTYLDSPVLD---GLAAL---AAQGAPT 266  
Db 969 VIHDPSPRCRELSGHAERALRAAGASPVLPVDAPAADRASFAALLRSATGPD 1022

## RESULT 2

S-09-105-537-2  
; Sequence 2, Application US/09105537A  
; Patent No. 6265202

GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.  
APPLICANT: Xue, Y.

APPLICANT: Zhao, L.  
FILE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A

CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2

LENGTH: 5215  
TYPE: PRT

ORGANISM: Streptomyces venezuelae  
US-09-105-537-2

Query Match 7.5%; Score 105; DB 4; Length 5215;  
Best Local Similarity 24.6%; Pred. No. 0.14;  
Matches 87; Conservative 24; Mismatches 105; Indels 138; Gaps 20;  
QY 36 LDLD-----ALGRWLATLP-----RRPQAL-----GNNV 60  
Db 3473 LDLDACALVAARGRLMQLPPGGAMSVRAGEDEVALLAGREDAVCAVAVNGPRSVVI 3532  
61 AGLARGEIAATLRAG-CCDIRWLEA-----QPLAMGLNGYR-----97  
Db 3533 SGAEEAEEAQAAGRRTRRLRVAAHAFHSPMDMLAGFVEAAGLRYPETLVVS 3592  
QY 98 -----NPDOL-GADRWACMGVILARQP-----SVHPDLLVASFGTATTLTIGPD 141  
Db 3593 TVTGRPARPGELTGDYWAQV-----REPVRFADAVRTAH-----RLGARTFLET-GPD 3641  
QY 142 NVF-----PGGLILPGPAMRGALAYGTAHLPLADGLVAD 176  
Db 3642 GVLGMAEECEDDDTVALLPAHKPGTAPHGPAAPG-ALRAAAAYGRGARVDWAGMHAD 3700  
QY 177 YPIDTHOAIASGIAAQAAGAIWOMLA-CR-----ORYGOAPEIYVAGGG-----WP 222  
Db 3701 GPEGPARVELPVHAFRHR---RYWLAPGAAATDDWYRIGWDRPLPATVGGARTAGRWL 3757  
QY 223 EVROEAERLLAVTG-----AAGFATPQPTYLDSPVLD---GLAAL---AAQGAPT 266  
Db 3758 VIHDPSPRCRELSGHAERALRAAGASPVLPVDAPAADRASFAALLRSATGPD 3811

## RESULT 3

5310667-9

Patent No. 5310667

APPLICANT: EICHHOLTZ, DAVID A.; GASSER, CHARLES S.; KISHORE, GANESH M.

TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL-3-PHOSPHOSHIKIMATE SYNTHASES

NUMBER OF SEQUENCES: 37

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/380,963

FILING DATE: 17-JUL-1989

SEQ ID NO: 9

LENGTH: 440

5310667-9

Query Match 7.4%; Score 102.5; DB 6; Length 440;

Best Local Similarity 26.2%; Pred. No. 0.0074;

Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQORALGVNVAGLAR-----GEATAATIRAGGCCDIR-----81

Db 115 LSGVPRMHERPIGDLVDALRQFNAGIEYLQAGYPPPLRIGGGSIRVDGPVRVSGSVSQF 174

QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRWACMGVILARQPSVHPPL-LVASFGTATTLT- 136

Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPIEITLNLMARFGVSVRRDG 225

QY 137 ---TIGPDNVE--PGGLILPGPAMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183

Db 226 WRAFTIARDAYVRGPGRMAIEGDA-----STASYFLALGAIGGGPVRTGVGVEDSIQ 277

QY 184 A-IASGIAAQAAGAIWROWLAGRQRYGOAPEIYVAGGWPVEV---QEAERLLA----- 233

Db 278 GDVAFEAATLAANGADVRY-----GPGWIEIRGVRVAEGRLKAPDADF 320

QY 234 ---VTGAAGFATPQPTYLDSP 251

Db 321 NLIPDAAMTAATLALYADGP 340

## RESULT 4

US-08-476-008-64

Sequence 64, Application US/08476008

Patent No. 5627061

GENERAL INFORMATION:

APPLICANT: Barry, Gerard F.

APPLICANT: Kishore, Ganesh M.

APPLICANT: Padgett, Stephen R.

APPLICANT: Stallings, William C.

TITLE OF INVENTION: Glyphosate Tolerant

TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F

STREET: 700 Chesterfield Village Parkway

CITY: St. Louis

STATE: Missouri

COUNTRY: USA

ZIP: 63198

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,008

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/306,063

FILING DATE: 13-SEP-1994

APPLICATION NUMBER: US 07/749,611

FILING DATE: 28-AUG-1991

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/576,537

;; FILING DATE: 31-AUG-1990  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hoerner Jr., Dennis R.  
;; REGISTRATION NUMBER: 30,914  
;; REFERENCE/DOCKET NUMBER: 38-21(10660)A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (314)537-6099  
;; TELEFAX: (314)537-6047  
;; INFORMATION FOR SEQ ID NO: 64:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 442 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-476-008-64

Query Match 7.4%; Score 102.5; DB 1; Length 442;  
Best Local Similarity 26.2%; Pred. No. 0.0075;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQALGVNVAGLAR-----GEATAATLRAGGCDIR-----81  
Db 115 LSGVPRMHERPIGDLVLRQFGAGIEYLQAGYPPLRIGGSIKRVDPVVRVGVSSQF 174  
QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRACWACVGLARQPSVHPPL-LVASFGTATLTD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYEITLNLMAFGVSVRRDG 225  
QY 137 ----TIGPDNVF--PGGLILPGPMRMGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAFTIARDVYRGPGMAIEGDA-----STASYFLALGAIGGGPVRTVGVEDSIQ 277  
QY 184 A-IASGIAAQAQAIIVRWLAGRQRYGQAPETIYVAGGWPVVR-----QEAERLLA----- 233  
Db 278 GDVAFATLAAMGADVRY-----GPGWETRGVRVRAEGGRLKAFDADF 320  
QY 234 --VTGAAGFATPQPTLYDSP 251  
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 5  
US-08-306-063-64  
; Sequence 64, Application US/08306063  
; Patent No. 5633435  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glyphosate Tolerant  
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/306,063  
; FILING DATE: 13-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/749,611  
; FILING DATE: 28-AUG-1991

;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/576,537  
;; FILING DATE: 31-AUG-1990  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hoerner Jr., Dennis R.  
;; REGISTRATION NUMBER: 30,914  
;; REFERENCE/DOCKET NUMBER: 38-21(10660)A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (314)537-6099  
;; TELEFAX: (314)537-6047  
;; INFORMATION FOR SEQ ID NO: 64:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 442 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-306-063-64

Query Match 7.4%; Score 102.5; DB 1; Length 442;  
Best Local Similarity 26.2%; Pred. No. 0.0075;  
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQALGVNVAGLAR-----GEATAATLRAGGCDIR-----81  
Db 115 LSGVPRMHERPIGDLVLRQFGAGIEYLQAGYPPLRIGGSIKRVDPVVRVGVSSQF 174  
QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRACWACVGLARQPSVHPPL-LVASFGTATLTD- 136  
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYEITLNLMAFGVSVRRDG 225  
QY 137 ----TIGPDNVF--PGGLILPGPMRMGALAYGTAHLPLADGLVADYPI-----DTHQ 183  
Db 226 WRAFTIARDVYRGPGMAIEGDA-----STASYFLALGAIGGGPVRTVGVEDSIQ 277  
QY 184 A-IASGIAAQAQAIIVRWLAGRQRYGQAPETIYVAGGWPVVR-----QEAERLLA----- 233  
Db 278 GDVAFATLAAMGADVRY-----GPGWETRGVRVRAEGGRLKAFDADF 320  
QY 234 --VTGAAGFATPQPTLYDSP 251  
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 6  
US-08-833-485-64  
; Sequence 64, Application US/08833485  
; Patent No. 5804425  
; GENERAL INFORMATION:  
; APPLICANT: Barry, Gerard F.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Padgett, Stephen R.  
; APPLICANT: Stallings, William C.  
; TITLE OF INVENTION: Glyphosate Tolerant  
; TITLE OF INVENTION: 5-Enolpyruvylshikimate-3-Phosphate Synthases  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Village Parkway  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/833,485  
; FILING DATE: 07-APR-1997  
; CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,063
; FILING DATE: 13-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,611
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/576,537
; FILING DATE: 31-AUG-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoerner Jr., Dennis R.
; REGISTRATION NUMBER: 30,914
; REFERENCE/DOCKET NUMBER: 38-21(15117)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)737-6099
; TELEFAX: (314)737-6047
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 442 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-833-485-64

Query Match 7.4%; Score 102.5; DB 1; Length 442;
Best Local Similarity 26.2%; Pred. No. 0.0075;
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQALGVNVAGLAR-----GEAATLRAAGCGDIR----- 81
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 LSGVPRMERPIGDLDALRQFGAGIEYLGQAGYPLRIGGGSIRVDGPVRVEGVSQF 174
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRWACMGVGLARQPSVHPPL-LVASFGTATLTD- 136
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYEITLNLMARFGVSRRDG 225
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 ---TIGPDNVF--PGGLILPGPAMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 WRAFTIARDAVYRGPGRMAIEGDA-----STASYFLALGAIGGGPVRVTGVGDSIQ 277
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 A-IASGIAAQAQAVRWLQAGRQRYGQAPETIYVAGGWEVPR-----QEAERLLA----- 233
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 GDVAFATLAAMGADVRY-----GPGWIETRGVRAEGGRLKAFDAF 320
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 --VTGAFAFCATPQPTYLDSP 251
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 7
US-09-137-440-64
; Sequence 64, Application US/09137440
; Patent No. 6248876
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Padgett, Stephen R.
; APPLICANT: Stallings, William C.
; TITLE OF INVENTION: Glyphosate Tolerant
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF
; STREET: 700 Chesterfield Village Parkway
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,063
; FILING DATE: 13-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/749,611
; FILING DATE: 28-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/576,537
; FILING DATE: 31-AUG-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoerner Jr., Dennis R.
; REGISTRATION NUMBER: 30,914
; REFERENCE/DOCKET NUMBER: 38-21(15117)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)737-6099
; TELEFAX: (314)737-6047
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 442 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-833-485-64

Query Match 7.4%; Score 102.5; DB 4; Length 442;
Best Local Similarity 26.2%; Pred. No. 0.0075;
Matches 68; Conservative 22; Mismatches 83; Indels 87; Gaps 13;

QY 45 LATLPRRQALGVNVAGLAR-----GEAATLRAAGCGDIR----- 81
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 LSGVPRMERPIGDLDALRQFGAGIEYLGQAGYPLRIGGGSIRVDGPVRVEGVSQF 174
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 ---WLRAQPLAMGLRNGYRNPDLGADRWACMGVGLARQPSVHPPL-LVASFGTATLTD- 136
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 LTALLMAAPV-LARRS-----GQDITIEVVGELISKPYEITLNLMARFGVSRRDG 225
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 ---TIGPDNVF--PGGLILPGPAMRGALAYGTAHLPLADGLVADYPI-----DTHQ 183
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 WRAFTIARDAVYRGPGRMAIEGDA-----STASYFLALGAIGGGPVRVTGVGDSIQ 277
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 A-IASGIAAQAQAVRWLQAGRQRYGQAPETIYVAGGWEVPR-----QEAERLLA----- 233
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 GDVAFATLAAMGADVRY-----GPGWIETRGVRAEGGRLKAFDAF 320
; : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 --VTGAFAFCATPQPTYLDSP 251
; : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 NLIPDAAMTAATLALYADGP 340

RESULT 8
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
```

STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1996 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-9

Query Match 7.4%; Score 102.5; DB 2; Length 1996;  
Best Local Similarity 23.8%; Pred. No. 0.066;  
Matches 94; Conservative 21; Mismatches 117; Indels 163; Gaps 18;

QY 4 LIDSGNSR--LKVGWFDPPDAP-----QAAREPAPVAFDNLDDALGRWLATLP-- 49  
DB 1544 LVDRHKVHLLVGRGGPDAPGVDRVLAELTESGAEVAVRACDVTDRLRLLDALPDE 1603  
QY 50 -----RRPQALGVNVAGLARG-----EAIAA 71  
DB 1604 HPLTCVVHTAGVLDGVLSSAQTAEIDTVLRPKADAHVHDELTRTGRLVPLVLYSSVA 1663  
QY 72 TLRAGGCDIRWLRAQPLAMGLRNGYRNP-----DOLGADRWACMVGVLARQPSVHPPLLV- 126  
DB 1664 TLGSAG-----QAGYAAANAFMDALAARR--CAAG-----HPALSIG 1698  
QY 127 ----ASFGTATTIDTIGPD-----NVFPGGLILPGPMMRGAL----- 160  
DB 1699 WGMWSGVGLATGLD--GADAAVRVRSGLAGPLDAGAALDLDLDRALTRPEALLPVRLDLRA 1756  
QY 161 AYTGAHLPLADGLVADYPIDTHQ-----AIASGIAAAQAGAVIROWLAGROR 207  
DB 1757 AGATALPEVRLDLAGVPADARSTPGAAGTGDGAVRPAPAPADAAGTLAARLAGR-- 1814  
QY 208 YGOAPE-----IYVAGGGWPEVRQEA-----ERLLAVTGA 237  
DB 1815 --SAPERTALLDLVRTEVAAVLGHGDPAAIGAARTEKDGAFDLSLTAVDLNRNLTGTGL 1872  
QY 238 AFGAT-----POPTYLDSPLVDGLAALAAQAGPTA 267  
DB 1873 RLPATLVFDHPTPLALAEILLDGLA-AGPAEPAA 1906

RESULT 9  
US-08-804-198-3  
; Sequence 3, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138

STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1996 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-3

Query Match 7.4%; Score 102.5; DB 2; Length 1996;  
Best Local Similarity 23.8%; Pred. No. 0.066;  
Matches 94; Conservative 21; Mismatches 117; Indels 163; Gaps 18;

QY 4 LIDSGNSR--LKVGWFDPPDAP-----QAAREPAPVAFDNLDDALGRWLATLP-- 49  
DB 1544 LVDRHKVHLLVGRGGPDAPGVDRVLAELTESGAEVAVRACDVTDRLRLLDALPDE 1603  
QY 50 -----RRPQALGVNVAGLARG-----EAIAA 71  
DB 1604 HPLTCVVHTAGVLDGVLSSAQTAEIDTVLRPKADAHVHDELTRTGRLVPLVLYSSVA 1663  
QY 72 TLRAGGCDIRWLRAQPLAMGLRNGYRNP-----DOLGADRWACMVGVLARQPSVHPPLLV- 126  
DB 1664 TLGSAG-----QAGYAAANAFMDALAARR--CAAG-----HPALSIG 1698  
QY 127 ----ASFGTATTIDTIGPD-----NVFPGGLILPGPMMRGAL----- 160  
DB 1699 WGMWSGVGLATGLD--GADAAVRVRSGLAGPLDAGAALDLDLDRALTRPEALLPVRLDLRA 1756  
QY 161 AYTGAHLPLADGLVADYPIDTHQ-----AIASGIAAAQAGAVIROWLAGROR 207  
DB 1757 AGATALPEVRLDLAGVPADARSTPGAAGTGDGAVRPAPAPADAAGTLAARLAGR-- 1814  
QY 208 YGOAPE-----IYVAGGGWPEVRQEA-----ERLLAVTGA 237  
DB 1815 --SAPERTALLDLVRTEVAAVLGHGDPAAIGAARTEKDGAFDLSLTAVDLNRNLTGTGL 1872  
QY 238 AFGAT-----POPTYLDSPLVDGLAALAAQAGPTA 267  
DB 1873 RLPATLVFDHPTPLALAEILLDGLA-AGPAEPAA 1906

RESULT 10  
US-08-804-227C-2  
; Sequence 2, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4472 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-2

Query Match 7.2%; Score 100; DB 2; Length 4472;  
Best Local Similarity 24.2%; Pred. No. 0.4;  
Matches 72; Conservative 24; Mismatches 107; Indels 94; Gaps 13;  
QY 39 DALGRWLATLPRRQALGVNVLARGEAIAATLRAGC-----GCDIRWLRAQP 87  
DB 3602 DESGRRVVAHSAPEAAVHSAEG---GDSAGVWTRHGEGTLVPDPEPTPPDADWARAWP 3658  
QY 88 LA-----MGLRNGYRNPDLGA-----DRWACHVGVLARQP 118  
DB 3659 PAGERVEPAELYERFAGLGEYGEAFAGVAVRWQPDALLAEVLLPDRASTGAG-----RF 3714  
QY 119 SVHPPLLVASFATTLDTIGPDNVPFGGLI-LPGPAMM-----RGALAYGTAHLPL---- 169  
DB 3715 GVHPALLDAAL-----QPWTAGLLEVPEDAVLLPFAWQGVSLYATGAGALRYR 3763  
QY 170 ----ADGLVADYPIITHQAIAGSIAAQAIVROWLAGRQ-----RYGQ 210  
DB 3764 LTKAGDGAVSLQAADT-----SGAAVLSLGLAVMRPLARRKLDVLLGTGAGERSLYRVEW 3818  
211 APEIYVAG--GGWPEVROEAERLLAVTGAAGATPQPTVLDSPVLDGLAALAAQAP 265  
DB 3819 QPRLLPAGPPRSWAVLGPADRLAGTPTGLGDPQDGGPTAL-YPEVRALKALAAQAP 3874

RESULT 11  
5310667-6  
Patent No. 5310667  
APPLICANT: EICHHOLTZ, DAVID A.; GASSER, CHARLES S.; KISHORE,  
GANESH M.  
TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL  
-3-PHOSPHOSHIKIMATE SYNTHASES  
NUMBER OF SEQUENCES: 37  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/380,963  
FILING DATE: 17-JUL-1989  
SEQ ID NO: 6;  
LENGTH: 382  
5310667-6

Query Match 6.9%; Score 95.5; DB 6; Length 442;  
Best Local Similarity 24.9%; Pred. No. 0.043;  
Matches 49; Conservative 29; Mismatches 72; Indels 47; Gaps 9;

QY 42 GRWLATLPRRQALGVNVLARGEAIAATLRAGC-----LRAQP-----LAMG 91  
DB 81 GIPASLDSKSDIELYGNAGTAMRPLTAATAVTAAGG-NARYVLDGVPMRERPGDLVVG 139  
QY 92 LRNGYRNPDLGADRWACHVGVVLARQPSVHPPLLVASFATTLDTIGPDNVPFGGLILP 151  
DB 140 LK-----QLGAD-VECTLTGTC-----PPVRVNAAG-----LRGKVL 173  
QY 152 G-----PAMRGALAYGTAHLPLADGLVADYPIITHQAIAS--GIAAQAQAIVROW 201  
DB 174 GSISSOYLTAALLMAAPLALGDVBIETIIDLISVPYVEMTLKMERFGVSAHSDSDWRFF 233  
QY 202 LAGRORYGOAPEIYVAG 218  
DB 234 VKGGQIKSPGNAYVEG 250

RESULT 12  
US-08-804-227C-4  
Sequence 4, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: DeHoff, Bradley S.  
APPLICANT: Kubstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3729 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-4

Query Match 6.7%; Score 93.5; DB 2; Length 3729;  
Best Local Similarity 24.1%; Pred. No. 1.6;  
Matches 84; Conservative 25; Mismatches 115; Indels 125; Gaps 20;  
QY 16 WFD-----PDAPQAAREPA-----PVAFNL-DLDAL-----GRWLATLPRR---P 52  
DB 930 WLDTTPTTPTTQSPDQAQNPADALFYKYSWRLRQDQSLARLDGRVLLVWPEASDP 989  
QY 53 QRALGVNVLARGEAIAATLRAGC-----AOPL-----AMGLRNGY 96  
DB 990 SVAEGVARELTANGATVESLTVEPGADRSRLRGLLVDTADTERDAGPGRGIVSLALAG-- 1047  
QY 97 RNPDLQAG-----RWACHVGVLARQPSVHP---PLL 125  
DB 1048 ---DHAGADGARPVPVAGLAASLALQAAGDAGTAEAGLWAVTRGAAVAVPGDVFPASOAL 1104

```
QY 126 VASFGTATLDTGPDNVEP---GGLI-LP-GP-----AMMERGAL 160
Db 1105 LWGFGVRVAGIE-----LPHCWGGLDLDTGPGDSGFRQLATLAGRPADQVALRASG 1157
QY 161 AYG--TAHLPLADGLVADYIDTHQATASGIAAQAAGAIYROWLAGRQRYGQAPFIYVAG 218
Db 1158 AYGRLVRAAAGAGDWRPRGT--VLVVGDTAEVAGPLVR-WLLG-----NGARRVTLG 1210
QY 219 GHPREVRQERLLAVTGAFAFGATPQPTYLDSPVLDGLAALAAQAFTA 267.
Db 1211 LSGPLPELADVAARVTAPCDPADRP-----ALRTLLAEQAFTA 1250

RESULT 13
US-08-374-483-2
; Sequence 2, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 820 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1831..4290
; OTHER INFORMATION: /note= "Xaa = Ter"
US-08-374-483-2

Query Match 6.5%; Score 91; DB 2; Length 820;
Best Local Similarity 23.6%; Pred. No. 0.33;
Matches 86; Conservative 27; Mismatches 127; Indels 124; Gaps 19;

QY 8 GNSRLKVGWFDPAQAAREPAPVAFDNLDDLALGRWLATLPRPQALGVNVAGLARGE 67
Db 464 GCCTAVXWHP--AQGRSYPHRFAPFTLXVQMGSGCLPAGEXRKRFPXGGRSAGKAGS 521
QY 68 AIAATLARGGCDIRWLAQPLAMGLRNGY-----RNPDLGADRWAC----- 109
Db 522 XAAATYRS-----RWARKSHLLPGATGXESCSCRHPXAGGLRACPLACFPXNPPE 576
QY 110 -----WVGVLARQPSV-----HPP----- 111

QY 577 GARPAIAVLARKQSFSTVXDRPPXACFXAFDQAVPGGPTARSAPALRHLDPAYLLVSRVG 636
QY 131 TA-----TTLDTTGPDNVFP-----GGLILPGPAMRGALAYGTA---HLP- 168
Db 537 AAFVAVRXSVLVQGTGQGHVFPRAQGRPRORSIGHGEGVRSGLRAGOGALEAGPAGALPV 696
QY 169 --LADGLVA--DYPIDTH--QAIASGIAAAQAAGAIYROWL-----AGRQRY- 208
Db 697 FALRVGVAFDHGVIVQPLRGVALGAQLALGGAARGAVQTFEGVELGREKYFRGVRGIR 756
QY 209 --GQAPET--YVAGGWP--EVRQEAERLL--AVTGAAF-----GATQPIY--LDS 250
Db 757 AAGPADGLAFHEPGELWPPFGVKQVSPMLFDFALTSGFHEPVSTLGDDEKAVRPVYRLER 816
QY 251 PVLD 254
Db 817 PVLE 820

RESULT 14
US-08-374-483-5
; Sequence 5, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 820 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4368..6827
US-08-374-483-5

Query Match 6.5%; Score 91; DB 2; Length 820;
Best Local Similarity 23.6%; Pred. No. 0.33;
Matches 86; Conservative 27; Mismatches 127; Indels 124; Gaps 19;

QY 8 GNSRLKVGWFDPAQAAREPAPVAFDNLDDLALGRWLATLPRPQALGVNVAGLARGE 67
Db 464 GCCTAVXWHP--AQGRSYPHRFAPFTLXVQMGSGCLPAGEXRKRFPXGGRSAGKAGS 521
QY 68 AIAATLARGGCDIRWLAQPLAMGLRNGY-----RNPDLGADRWAC----- 109
Db 522 XAAATYRS-----RWARKSHLLPGATGXESCSCRHPXAGGLRACPLACFPXNPPE 576
QY 110 -----WVGVLARQPSV-----HPP----- 111
```

Db 522 XAAATYRS-----RWAKSHLLPGATGXSXESCRHPXAGGPIRXACPLACFPXPNPPE 576  
QY 110 -----WVGVLARQPSV-----HPP-----LLVASFG 130  
Db 577 GARPAIAVIAARKQSFSTYXDRPPXACFXAFDQAVPGGPTARSPALRHLDPPAYLLVSRVG 636  
QY 131 TA-----TTLDTIGPDNVFP-----GGILPGPAMMRGALAYGTA---HLP- 168  
Db 637 AAFVQRQSVLQVTCGGHVPFRAQGPQRSLRGHGEVRSGLRAGOCALGAGPAGAEALPV 696  
QY 169 --LADGLVA-DYPIDTH--QAIASGIAAQAQAIIVROWL-----AGRQRY----- 208  
Db 697 FALRVGQVAFDHGVIVQPLRGVALGAQLALGGAARGAVQTEGVELGREKYFRFGVGIR 756  
QY 209 --GQAPEI--YVAGGWP-EVQREARLL--AVTGAAF-----GATPOPTY-LDS 250  
Db 757 AAGPADGLAFHPEGLMFPFGVKNQVSPMLFDAFLTSGFHEPVSTLGDKAKVRVPVYRLER 816  
QY 251 PVLD 254  
Db 817 PVLE 820

RESULT 15  
US-08-804-227C-3  
; Sequence 3, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: DeHoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostack, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: THOMAS G. PLANT 1501  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: ASCII(DOS) Text only  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,227C  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plant, Thomas, G.  
; REGISTRATION NUMBER: 35,784  
; REFERENCE/DOCKET NUMBER: X-8231  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-2459  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1864 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-804-227C-3

Query Match 6.5%; Score 91; DB 2; Length 1864;  
Best Local Similarity 25.0%; Pred.No.1.1;  
Matches 78; Conservative 29; Mismatches 123; Indels 82; Gaps 17;  
QY 2 IILIDSGNSRLKGVWFDPAQAPAPAPVAFDNLDDALGR---WLATLPRRQORAL-- 56  
Db 1022 LVLPEDEGEVRLRVLAEPDA-SGYRE-----LSVHSAGEDGWN---TFHATAVLDT 1068  
QY 57 GNVNAGLARGEIAATLRAGCCDINWLRQAQ-----PLAMGLRNGYRNPDPQLGA 104

Db 1069 GTTTAGEPAG-APPAWPPGDAEPDLAAEYERFADAGIGYGFAPFRLRSARWRCGDAILA 1127  
QY 105 DRWACMVGYLARQP-----SVHPPPLLVSFGTATTLDITIGPDNVFP---GGILPGPAMMRG 158  
Db 1128 D--VRLPGELAGEADRYGIHPALLDAALHTAAAA-LGGAHGMLPFTWNGVTLHA---RG 1180  
QY 159 ALAYGTAHLPLADGLVADYPID-----THQATIASG-IAAAQAQAIIVROWLAGR 205  
Db 1181 AHAIKRVRLTPAGPDVAVAVTAVDPAGRPVFTAAASLTLRPVTTGQLTAAEA----- 1229  
QY 206 QRYGOAPEIYVAGGWPVEVQREARLLAVTGAAF-----GATPOPTYLDSVLDG----- 255  
Db 1230 ---ARAPLYRVVWTGTLPTGTARDHTWAVAGGPGDLLPGETPHHPDLASALADTGTAPFR 1286  
QY 256 -LAALAAQAGAPT 266  
Db 1287 VLADLRGYGTAT 1298

Search completed: June 24, 2003, 21:49:31  
Job time : 9.7999 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 12.1733 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-15

Perfect score: 1392

Sequence: 1 MIILDSGNSRLKVGWFDPD.....LDSPLDGLAALAAQAQAPTA 267

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

--Searched: 417779 seqs, 108206813 residues

-otal number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	267	9	US-09-813-453A-15
2	258.5	18.6	460	9	US-09-813-453A-39
3	256.5	18.4	592	9	US-09-813-453A-22
4	254.5	18.3	592	9	US-09-813-453A-43
5	194.5	14.0	242	9	US-09-813-453A-65
6	178.5	12.8	248	9	US-09-813-453A-20
7	163	11.7	244	9	US-09-813-453A-41
8	163	11.7	249	9	US-09-813-453A-61
9	154.5	11.1	241	9	US-09-813-453A-63
10	152	10.9	246	9	US-09-813-453A-9
11	148	10.6	255	9	US-09-813-453A-7
12	141	10.1	250	9	US-09-813-453A-3
13	140.5	10.1	258	9	US-09-813-453A-49
14	138	9.9	212	9	US-09-813-453A-59
15	138	9.9	262	9	US-09-813-453A-45
16	136.5	9.8	257	9	US-09-813-453A-53
17	134.5	9.7	260	9	US-09-813-453A-51
18	134	9.6	258	9	US-09-813-453A-6
19	134	9.6	265	9	US-09-813-453A-4

20	132	9.5	249	9	US-09-813-453A-70
21	131.5	9.4	254	9	US-09-813-453A-47
22	130.5	9.4	219	9	US-09-813-453A-57
23	129.5	9.3	258	9	US-09-813-453A-2
24	127.5	9.2	256	9	US-09-813-453A-55
25	125	9.0	233	9	US-09-813-453A-17
26	105	7.5	5215	9	US-09-860-846-2
27	105	7.5	5215	9	US-09-988-384B-2
28	105	7.5	5215	9	US-09-836-821-2
29	105	7.5	5215	10	US-09-861-289-2
30	103.5	7.4	262	9	US-09-813-453A-8
31	102.5	7.4	442	9	US-09-464-099A-64
32	102.5	7.4	442	9	US-10-214-766-36
33	102.5	7.4	442	10	US-09-861-696-64
34	97	7.0	257	9	US-09-813-453A-13
35	97	7.0	562	9	US-09-712-363-254
36	95	6.8	408	9	US-09-738-626-5136
37	94.5	6.8	437	9	US-10-145-415-101
38	92.5	6.6	272	9	US-09-712-363-276
39	92.5	6.6	272	9	US-09-813-453A-5
40	90.5	6.5	891	9	US-09-976-059-18
41	85.5	6.4	259	10	US-09-820-893-87
42	89.5	6.4	442	10	US-09-749-728B-11
43	89	6.4	4150	9	US-09-808-880-2
44	88.5	6.4	190	10	US-09-764-864-1071
45	88.5	6.4	1023	9	US-09-893-519A-14

#### ALIGNMENTS

##### RESULT 1

US-09-813-453A-15  
; Sequence 15, Application US/09813453A  
; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; APPLICANT: Patterson, Thomas A.

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15

; LENGTH: 267

; TYPE: PRT

; ORGANISM: Bordetella pertussis

; US-09-813-453A-15

Query Match 100.0%; Score 1392; DB 9; Length 267;  
Best Local Similarity 100.0%; Pred. No. 3.2e-114;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIILDSGNSRLKVGWFDPDPAQAAREPAPVAFNLDLALGRWLATLPRPQALGVNV 60

1 MIILDSGNSRLKVGWFDPDPAQAAREPAPVAFNLDLALGRWLATLPRPQALGVNV 60

QY 61 AGLARGEATATIRAGGCDIRWLRAPOLGADRWACWGVVLARQPSV 120

61 AGLARGEATATIRAGGCDIRWLRAPOLGADRWACWGVVLARQPSV 120

QY 121 HPPLLVASFGTATLTDTIGPDNVFPGGLIILPGPAMRGALAYGTAHLPLADGLVADYDID 180

121 HPPLLVASFGTATLTDTIGPDNVFPGGLIILPGPAMRGALAYGTAHLPLADGLVADYDID 180

QY 181 THQAIASGIAAAGATVROWLAGRORYGOAPEYVAGGWPVEYRQAEERLLAVTGAAGF 240

181 THQAIASGIAAAGATVROWLAGRORYGOAPEYVAGGWPVEYRQAEERLLAVTGAAGF 240



Db 510 NAVASGMDDAVCGSYM--MMHGRKKEKTGAGKPDVVIITGGG-----AAKVAEALPPA 560  
QY 239 FGATPQPTYLDSPVLGDLA-ALAAQG 263  
Db 561 FLAENTVRVADNLVIVYGLLNMAAEG 586

RESULT 5

US-09-813-453A-65  
; Sequence 65, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 65  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-09-813-453A-65

Query Match 14.0%; Score 194.5; DB 9; Length 242;  
Best Local Similarity 27.9%; Pred. No. 1.8e-09;  
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;  
QY 4 LIDSGNSRLKVGWFDPAQOAREPA-----PVAFDNLDDLALGRWLATLPRRQAL 56  
Db 5 LFDLGNRFKA-----SLREGVIGPVTVLPYTETMDAFA---LQELPRG----- 47  
QY 57 GNVNVLARGEA-----IAATLRAGGDIRWLRAQPLAMGLRNGYR 97  
Db 48 --RVAVLASVAAPATHVLEVKIHFEQVQVAA-----CAGVRIAYA 92  
QY 98 NPQOLGADRWACWVGVLAROPSVHPPLLVASFATTTLDITGPDNVPFGGLILPGPAMMR 157  
Db 93 HPERFGVDFRLLALGYSYG-----EGNVLVVGVTALTIDLLAANGCHLGRISASPTLMR 147  
QY 158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAQAAGAIVRWLACRQRYGQAPETIYA 217  
Db 148 QALHARAEQLPLSGGNYLEPAEDTEDALVSGCNGAVALJERSLYEAHQRLDQSVRLHH 207  
QY 218 GGGWPEVROBAERLLAVTGAAGFACPTQPTVLDSPVLDGLAALAAQA 264  
Db 208 GGG-----VASLLPWLG---DVVHRPTL-----VLGLAIWAARAA 240

RESULT 6

US-09-813-453A-20  
; Sequence 20, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-813-453A-20

Query Match 12.8%; Score 178.5; DB 9; Length 248;  
Best Local Similarity 29.6%; Pred. No. 4.6e-08;  
Matches 75; Conservative 93; Mismatches 98; Indels 53; Gaps 11;  
QY 1 MIIIDSGNSRLKVGWFDPAQOAREPAVPVAFDNLDDLALGRWLAT---LPRRQALG 57  
Db 1 MILEDCGNSLIK--WRVIEG--AARSVAGGLAES--DDALVEQLTSSQALPVRACRLVS 54  
QY 58 VN-----VAGLARGEIAAATLRAGGDIRWLRAQPLAMGLRNGYRNPQOLGADRWAC 109  
Db 55 VRSEQETSQVLARLEQLFPVSALVASSGKQL-----AGVRNGYLDYQRLGLDRWLA 105  
QY 110 MVGV--LARQPSVHPPLLVASFATTTLDITGPDNVPFGGLILPGPAMMARGALAYGTAHL 167  
Db 106 LVAAHHLAKK-----ACLVIDLGTAVTSDLVAAADGVHLGGYICPGMTLMRSQRLTHTRRI 160  
QY 168 PLADGLVADYPIDTHQAIAS---GIAAAQA-----GAIVRWLACRQRYGQAPETI 214  
Db 161 RYDDA-----EARRALASLOPQONTAEAVERGCLLMRLGRFVREQYAMACELLGPDCEI 213  
QY 215 YVAGGGWPEVROE 227  
Db 214 FLTGDDAELVRDE 226

RESULT 7

US-09-813-453A-41  
; Sequence 41, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-09-813-453A-41

Query Match 11.7%; Score 163; DB 9; Length 244;  
Best Local Similarity 30.8%; Pred. No. 1e-06;  
Matches 56; Conservative 28; Mismatches 78; Indels 42; Gaps 10;  
QY 3 ILIDSGNSRLKVGW-----FDP-DAPQA-AREPAPVAFDNLDDLALGRWLATLPR 50  
Db 5 LIVDQGSACKVAEVRNNSIESISFLPGKAGQALSHLVAFHFRFOKAIYSSVG-----LPD 59  
QY 51 RPOQALGVNVLARGEIAAATLRAGGDIRWLRAQPLAMGLRNGYRNPQOLGADRWACM 110  
Db 60 EEAEAIVRSCA-----AASLMG-----TETVPVLRQLQVDRRT---LGAORLAIV 101  
QY 111 VGVLAROPSVHPPLLVASFATTTLDITGPDNVPFGGLILPGPAMMARGALAYGTAHLPLA 170  
Db 102 VGHSLYPNTE--LLVIDAGTAITYERSVAGSIIYGGNISPLGLHURFKALHFTORLPLI 159



QY 211 APEIYVAG 218  
 Db 209 LPVVLTTGG 216

## RESULT 11

US-09-813-453A-7  
 ; Sequence 7, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 255  
 ; TYPE: PRT  
 ; ORGANISM: Geobacter sulfurreducens  
 US-09-813-453A-7

Query Match 10.6%; Score 148; DB 9; Length 255;

Best Local Similarity 23.1%; Pred. No. 2.2e-05;  
 Matches 66; Conservative 46; Mismatches 110; Indels 64; Gaps 12;

QY 1 MIIIDSGNSRLKVGWFDPAQAPAPVAFDNLDDALGRW-LATLPRPQALGV- 58  
 Db 1 MLLVDVGNVILGIYDGE--RLVRD-----WRVSTDKARTDEYGI 42  
 QY 59 -----NVAGLARGEIAATLRA-----GGCDIRWLRAQPLAM--GLRNG----YRN 98  
 Db 43 INELFRLAGLGDQIRAVISSVVPLTGLVLERLSLGVFMRPLVVGPGIKTGMPIOYDN 102  
 QY 99 PDQLGADRWACWGVLARQPSVHPPLLVASFGTATTDTIGPDNVFPGGLILPGPAMRG 158  
 Db 103 PREVGADR---IVNAVAEYKRTSLIIVDFGTATTDFYVNRKGEYCGGAIAPGLVISTE 159  
 QY 159 ALAYGTAHLPLAD-----GLVADYPIDTHQAIASGIAAQAAGATVROWLAGRQRYGOAPE 213  
 Db 160 ALQFASKLPRVDIIRSAIARTVNSMQA---GIYGVGLVDEIVTRWKAESKDAPR 216  
 QY 214 IYVAGGWEVROEAEERLLAVTGAAGATPQTYLDSFVLDDGLAAL 259  
 Db 217 VIATGGLASLIAPESKTEIARV-----EEYL---TLEGLRIL 249

## RESULT 12

US-09-813-453A-3  
 ; Sequence 3, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3  
 ; LENGTH: 250  
 ; TYPE: PRT  
 ; ORGANISM: Clostridium acetobutylicum  
 US-09-813-453A-3

Query Match 10.1%; Score 141; DB 9; Length 250;

Best Local Similarity 22.5%; Pred. No. 8.8e-05;  
 Matches 58; Conservative 43; Mismatches 103; Indels 54; Gaps 8;

QY 1 MIIIDSGNSRLKVGWFDPAQAPAPVAFDNLDDALGRW-LATLPRPQALGVN 59  
 Db 18 VILVDVGNVILGIYN-----DTKLTAEWRLSTDVLSADEYGIO 59  
 QY 60 VAGLARGEIAATLRA-----GGCDIRWLRAQPLAM--GLRNG----YRN 98  
 Db 60 VMNLFQDQKLDPTLVEGVIISSVVPNIMYSLEHMIRKVFKNPLVVGPGIKTGINIKYDN 119  
 QY 99 PDQLGADRWACWGVLARQPSVHPPLLVASFGTATTDTIGPDNVFPGGLILPGPAMRG 158  
 Db 120 PKEVGADR---IVNAVAEYKRTSLIIVDFGTATTFCVAVRENGDYLGGAIAPGLVISTE 176  
 QY 159 ALAYGTAHLPLADGLVADYPIDTHQAIASGIAAQAAGATVROWLAGRQRYGOAPEIYVAG 218  
 Db 177 ALFEKAALKPRVELIKPAYAI-----CKNTISSIOSGIVYRYLRQVKYLFELKE----- 226  
 QY 219 GGWEPEVROEAEERLLAVTG 236  
 Db 227 -NLDPGRTRTSLVLTG 243

## RESULT 13

US-09-813-453A-49  
 ; Sequence 49, Application US/09813453A  
 ; Patent No. US20020168681A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yocum, R. Rogers  
 ; APPLICANT: Patterson, Thomas A.  
 ; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
 ; TITLE OF INVENTION: ANTIBIOTICS  
 ; FILE REFERENCE: OGZ-001  
 ; CURRENT FILING DATE: 2001-03-20  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR FILING DATE: 2000-09-21  
 ; NUMBER OF SEQ ID NOS: 77  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 49  
 ; LENGTH: 258  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus stearothermophilus  
 US-09-813-453A-49

Query Match 10.1%; Score 140.5; DB 9; Length 258;

Best Local Similarity 24.7%; Pred. No. 0.0001;  
 Matches 66; Conservative 42; Mismatches 106; Indels 53; Gaps 12;

QY 1 MIIIDSGNSRLKVGWFDPD-----APQAREPAPVAFDNLDDALGRWLATLPRPQOR 54  
 Db 1 MIFVDVGNVILGIYDGDDELKHHWRIETSKSTE-----DEYGMKAL----- 46  
 QY 55 ALGVNVAGL-----ARGEIAATL-----RAGGCDIRWLRAQPLAM--GLRNG----YRN 99  
 Db 47 ---LNHVGLQFSDINGIISSVVPIMPALERMKLYKHIFLVGPGIKTGIDKYNP 103  
 QY 100 DQLGADRWACWGVLARQPSVHPPLLVASFGTATTDTIGPDNVFPGGLILPGPAMRG 159  
 Db 104 REVGADR---IVNAVAGIHLXGSLIIVDFGTATTFCYINERKQYMGGAIAFGIMISTEA 160  
 QY 160 LAYGTAHLPLADGLVADYPI--DTHQAIASGIAAQAAGATVROWLAGRQRYGO--APEIY 215

Db 161 LFARAAKLPRIEIARDDIIGKNTVSAMQAGILYGVGV--EGIVSRMKAKSKIPPKVI 218  
QY 216 VAGGHWPEVRQEAERL-----LAVTG 236  
Db 219 ATGGLAPLASESDIIVDVPFLTFRG 245

## RESULT 14

US-09-813-453A-59  
; Sequence 59, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 59  
; LENGTH: 212  
; TYPE: PRT  
; ORGANISM: Desulfovibrio vulgaris  
US-09-813-453A-59

Query Match 9.9%; Score 138; DB 9; Length 212;  
Best Local Similarity 27.4%; Pred. No. 0.00013;  
Matches 62; Conservative 34; Mismatches 78; Indels 52; Gaps 12;  
QY 3 ILIDSGNSRLKVGWFDPPAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQALGVNV-- 60  
Db 6 LLFDIGTNTVKIG-----IATVAVTLSY--VLPTDPGQTTDSIGRLLE 48  
QY 61 ----AGLAR--GEATAATLRAG-----GCDIRWLRAQ-----PLAMGLRNGYRNP 100  
Db 49 VLRHAGLGPADVGACVASSWPVGNPLIRRRACE-RYLKRKLLFAPGDIAIPLDNRVERPA 107  
QY 101 QLGADRWACWGVGLARQPSVHPPLLVASFGTATLDTGPDNVPFGGLILPQPMAMRGA 159  
Db 108 EVGADR--LVAAYAAHRLYPGRSLVSDFGTATTFDCV-EGGAYILGLICPGVLSSAGA 164  
QY 160 LAYGTAHLPL-LADGLVADYPIIDTHQAIAS-----GIAAAQAGAI 197  
165 LSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFIQFPAAMTEGVL 210

## RESULT 15

US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT

; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 9.9%; Score 138; DB 9; Length 262;  
Best Local Similarity 23.7%; Pred. No. 0.00017;  
Matches 62; Conservative 46; Mismatches 100; Indels 54; Gaps 11;  
QY 1 MIILIDSGNSRLKVGWFDPPAPQAAREPAPVAFDNLDLDALGRW-LATLPRRPQALGVN 59  
Db 1 MIFVLDVGNNTNAVIGVFE-----EGELRQHRWMTDRHKTEDEYGM 42  
QY 60 VAGLARGEATAATLRAG-----GCDIRWLRAQPLAM--GLRNG-----YR 97  
Db 43 VKOLLEHEGLSPEDVKGIIIVSSVPPIMFALERMCE-KYFKIRPLVVGPGIKTGLNIKYE 101  
QY 98 NPDQLGADRWACWGVGLARQPSVHPPLLVASFGTATLDTGPDNVPFGGLILPQPMAMR 157  
Db 102 NPREGADR---IVNAVAGIHLVGSPLIIVDFGTATTTCYINEKHVMGGVITPGIMISA 158  
QY 158 GALAYGTAHLPLADGLVADYPI--DTHQAIASGIA---AAQAGAIVROWLAGRORYGOAP 212  
Db 159 EALYSRAAKLPRIETPKPSSVVGKNTVMSAMOSGILYGVGVGIVKRM---KEEAKOEP 215  
QY 213 EIVVAGGHWPEVRQEAERLIAV 234  
Db 216 KV-IATGGLAKLISEESNVIDV 236

Search completed: June 24, 2003, 22:24:03  
Job time : 13.1733 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 27.4124 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-15  
Perfect score: 1392  
Sequence: 1 MIILDSGNRLXVGFDPD.....LDSPVLGLAALAAQAQPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues 908470  
total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1392	100.0	267	23 AAU91162	Bordetella pertussis
2	258.5	18.6	455	21 AAU91162	Neisseria gonorrhoe
3	258.5	18.6	455	20 AAU91162	Neisseria gonorrhoe
4	258.5	18.6	460	23 AAU91167	Pantothenate kinas
5	258.5	18.6	592	20 AAU91168	Neisseria gonorrhoe
6	258.5	18.6	592	21 AAU91168	Neisseria gonorrhoe
7	256.5	18.4	455	21 AAU91168	Neisseria meningit
8	256.5	18.4	592	20 AAU91168	Neisseria meningit
9	256.5	18.4	592	21 AAU91168	Neisseria meningit
10	256.5	18.4	592	23 AAU91166	Pantothenate kinas

11	254.5	18.3	592	20 AAU91165	Neisseria meningit
12	254.5	18.3	592	21 AAU91165	Neisseria meningit
13	254.5	18.3	592	23 AAU91169	Pantothenate kinas
14	244	17.5	389	21 AAU91169	Neisseria meningit
15	194.5	14.0	242	23 AAU91180	Pantothenate kinas
16	186	13.4	189	20 AAU91180	Neisseria meningit
17	178.5	12.8	248	23 AAU91164	Pantothenate kinas
18	163	11.7	244	23 AAU91168	Pantothenate kinas
19	163	11.7	249	23 AAU91178	Pantothenate kinas
20	161	11.6	259	23 ABBA47661	Listeria monocytog
21	154.5	11.1	241	23 AAU91179	Pantothenate kinas
22	152	10.9	246	23 AAU91156	Thermotoga maritim
23	148	10.6	255	23 AAU91154	Geobacter sulfurre
24	141	10.1	250	23 AAU91150	Clostridium acetob
25	140.5	10.1	258	23 AAU91172	Pantothenate kinas
26	138	9.9	212	23 AAU91177	Pantothenate kinas
27	138	9.9	262	23 AAU91170	Pantothenate kinas
28	136.5	9.8	257	23 AAU91174	Pantothenate kinas
29	134.5	9.7	260	23 AAU91173	Pantothenate kinas
30	134	9.6	258	23 AAU91153	Rhodobacter capsul
31	134	9.6	265	23 AAU91151	Streptomyces coeli
32	132	9.5	249	23 AAU91182	Pantothenate kinas
33	131.5	9.4	254	23 AAU91171	Pantothenate kinas
34	130.5	9.4	219	23 AAU91176	Pantothenate kinas
35	129.5	9.3	258	22 AAU01243	B. subtilis novel
36	129.5	9.3	258	23 AAU91149	Bacillus subtilis
37	127.5	9.2	256	23 AAU91175	Pantothenate kinas
38	125	9.0	233	23 AAU91163	Pantothenate kinas
39	121.5	8.7	548	22 AAU49918	Propionibacterium
40	112	8.0	9477	22 AAU01044	Streptomyces nous
41	105	7.5	1841	18 AAU22605	Tylosin synthase
42	105	7.5	4630	21 AAU77177	S. venezuelae vep
43	103.5	7.4	262	23 AAU91155	Deinococcus radiop
44	102.5	7.4	1996	18 AAU22607	Platenolide syntha
45	102.5	7.4	1996	18 AAU23717	Platenolide syntha

## ALIGNMENTS

RESULT 1  
AAU91162  
ID AAU91162 standard; Protein; 267 AA.  
AC AAU91162;  
XX  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Bordetella pertussis pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
XX Bordetella pertussis.  
XX  
XX WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
XX 24-AUG-2001; 2001WO-US26531.  
XX  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RE, Patterson TA;  
XX  
XX WPI; 2002-269358/31.  
XX  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the coax protein -

[illegible]

RESULT 2  
AAAY38617  
ID AAAY38617 standard; Protein: 455 AA.

XX	AA	Y38617;	
XX	AC		
XX			
DT	08-OCT-1999	(first entry)	

Neisseria gonorrhoeae antigen encoded by partial ORF61.

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; treatment. Neisseria infection; meningitis. *Neisseria meningitidis*; *Neisseria gonorrhoeae*.

XX treatment, *Neisseria* infection; meningitis; septicaemia; gonorrhea.  
XX  
OS *Neisseria gonorrhoeae*.

XX PN W09924578-A2.

XX  
PD  
20-MAY-1999.  
YY

09-OCT-1998; 98WO-IB01665.  
PF  
XX

PR	01-SEP-1998;	98GB-0019016.
PR	06-NOV-1997;	97GB-0023516.
PR	06-NOV-1997;	97GB-0023516.
PR	06-NOV-1997;	97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997. 97GB-0025158.

PR 27-NOV-1997; 97GB-0002350;  
PR 10-DEC-1997; 97GB-0026147;  
PR 14-JAN-1998; 98GB-0000759;

XX  
PA  
v  
v  
(CHIR-) CHIRON SPA.

PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;

```

XX WPI; 1999-327407/27.
XX
XX
XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX
XX Claim 4; Page 175; 524pp; English.
XX
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX CC reading frames (ORFs) AA11972-212358. The antigenic proteins,
XX CC their fragments, their nucleic acids and antibodies are used for
XX CC diagnosis, prevention (as vaccines) or treatment of Neisseria
XX CC infections, such as meningitis, septicaemia and gonorrhea. Both
XX CC organisms are closely related. Fragments of the nucleic acids
XX CC are useful as hybridisation probes and antisense reagents.
XX
XX SQ Sequence 455 AA;
Query Match 18.6%; Score 258.5; DB 20; Length 455;
Best Local Similarity 31.8%; Pred. No. 1.6e-18;
Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;
QY 3 ILIDGNSRLKVGWFPDPAPQAREPAPVAFNDLDDALGRWLATLPRPQALGVNVA 62
DB 204 LLEGGNSRLKWAWE-NGTFTATVGSAPYR---DLSPLGAWEAKDGNVRIVCAVC 258
QY 63 LARGEATAATLRAGGCDIWLRAOPLMGLRGYRNPDOLGADRWACVMGVLARQPSVHP 122
DB 259 ESKKAQVKQLAR--KIEWLPSSAQALGIRNHYRHPHGSDRWFNALG--SRFRS-RN 312
QY 123 PLLVASGFGTATLDTIGPDNVNFPFGILLPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182
DB 313 ACVWVSCGTAVTVDALTDDGHLGGTIMPGFHLMKESLAVRTANLNRPAKRYPPPTTTG 372
QY 183 QATASGIAAQAQAVYRWQ--LAGRORYGQAPETIYVAGSGWPEVQRQERLLAVTGAAGF 240
DB 373 NAVASGMDADVCGSIMMHGRLEKNGACKPVDVITGG-----AAKVAEALPPAFL 425
QY 241 ATPQPTYLDSPLVDGLAAL-AAQG 263
DB 426 AENTVRVADNLVTHGLNLIAAEG 449

```

RESULT 3		
AAAY74908		
ID	AAAY74908 standard; Protein; 455 AA.	
XX		
XX	AAAY74908;	
XX		
XX	21-MAR-2000 (first entry)	
DT		
XX		
DE	Neisseria gonorrhoeae ORF 311 protein sequence SEQ ID NO:1290.	
XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;	
KW	antibacterial; gene therapy.	
XX		
OS	Neisseria gonorrhoeae.	
XX		
PN	WO9957280-A2.	
XX		
PD	11-NOV-1999.	
XX		
PF	30-APR-1999; 99WO-US09346..	
XX		
PR	01-MAY-1998; 98US-0083758.	
PR	31-JUL-1998; 98US-0094869.	
PR	02-SEP-1998; 98US-0098994.	
PR	02-SEP-1998; 98US-0099062.	
PR	09-OCT-1998; 98US-0103749.	
PR	09-OCT-1998; 98US-0103794.	
PR	09-OCT-1998; 98US-0103796.	



PR 25-FEB-1999; 99US-0121528.  
 XX (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI: 2000-062150/05.  
 DR N-PSDB; AA253670.  
 XX  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 vaccines and diagnostics  
 XX  
 PS Claim 2; Page 709; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254616 to AA254617  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254617 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 455 AA;  
 Query Match 18.6%; Score 258.5; DB 21; Length 455;  
 Best Local Similarity 31.8%; Pred. No. 1.6e-18;  
 Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;  
 QY 3 ILIDSGNSRLKGVNFDPAQAPAPVAFDNLDLALGRWLATLPRRQALGVNVAG 62  
 DB 204 LLEGGNSRLKAWVE-NGTFATVGSAPYR-----DLSPLGAWEAEKADGNVRIVGCAVCG 258  
 QY 63 LARGEATAATLRAGGCDIRWLRAPQLAMGLRNGYRNPDLGADRWACWVGVLARQPSVHP 122  
 DB 259 ESCKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPHEHSGDRWFNALG--SRFSS-RN 312  
 QY 123 PLLVASEGTATTLDTIGPDNVFPFGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182  
 DB 313 ACVVVSGCTAVTVDALTDGHLGTTIMPGFHLKESLAVRTANLNRPAKRYPPFTTGT 372  
 QY 183 QAIASGTAAGAGAIVRQW--LAGRQRYGQAPETIVAGGWPVEVQAEERLLAVTGAAG 240  
 DB 373 NAVASGMDAVCGSITMMHGRLEKNGAGKPVVDVITGGG-----AAKVAEALPPAFL 425  
 QY 241 ATPQTYLDSFVLDDGLAAL-AAQG 263  
 DB 426 AENTVRVADNLVHGLNLIAAEG 449  
 RESULT 4  
 AAU91167  
 XX AAU91167 standard; Protein; 460 AA.  
 XX  
 AC AAU91167;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Pantothenate kinase (Coax) #5.  
 XX  
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX

PN WO200216601-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 24-AUG-2001; 2001WO-US26531.  
 XX  
 PR 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX  
 PI Yocum RR, Patterson TA;  
 PI WPI: 2002-269358/31.  
 DR N-PSDB; ABK54188.  
 XX  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 PS Claim 8; Page 92-94; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 460 AA;  
 Query Match 18.6%; Score 258.5; DB 23; Length 460;  
 Best Local Similarity 31.8%; Pred. No. 1.6e-18;  
 Matches 84; Conservative 43; Mismatches 116; Indels 21; Gaps 8;  
 QY 3 ILIDSGNSRLKGVNFDPAQAPAPVAFDNLDLALGRWLATLPRRQALGVNVAG 62  
 DB 209 LLEGGNSRLKAWVE-NGTFATVGSAPYR-----DLSPLGAWEAEKADGNVRIVGCAVCG 263  
 QY 63 LARGEATAATLRAGGCDIRWLRAPQLAMGLRNGYRNPDLGADRWACWVGVLARQPSVHP 122  
 DB 264 ESCKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPHEHSGDRWFNALG--SRFSS-RN 317  
 QY 123 PLLVASEGTATTLDTIGPDNVFPFGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH 182  
 DB 318 ACVVVSGCTAVTVDALTDGHLGTTIMPGFHLKESLAVRTANLNRPAKRYPPFTTGT 377  
 QY 183 QAIASGTAAGAGAIVRQW--LAGRQRYGQAPETIVAGGWPVEVQAEERLLAVTGAAG 240  
 DB 378 NAVASGMDAVCGSITMMHGRLEKNGAGKPVVDVITGGG-----AAKVAEALPPAFL 430  
 QY 241 ATPQTYLDSFVLDDGLAAL-AAQG 263  
 DB 431 AENTVRVADNLVHGLNLIAAEG 454  
 RESULT 5  
 AAY38618  
 ID AAY38618 standard; Protein; 592 AA.  
 XX  
 AC AAY38618;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE Neisseria gonorrhoeae antigenic protein encoded by ORF61.  
 DE  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW treatment; Neisseria infection; meningitis; septicemia; gonorrhea.  
 XX



Db	204	LLDGGNSRLKWAWE-NGTFATVGSAPYR-----DLSPGAEWAEEKVDGNVRIIVGCAVCG	255
Qy	63	LARGEATAATLRAGCGDIRWLRAOPLAMGLRNGYRNPDOLGADRWACMGVGLAROPSVHP	122
Db	259	EFKAAQVEQLAR---KIEWLPSSNQALGIRNHYRHPPEHSGSDRFNALG--SRFS-RN	312
Qy	123	PLLVASFGTATLTDTIGPDNVFPGLILPGPAMRGALAYGTAHLPLADGLVADYPIDTH	182
Db	313	ACVVVSGCTAVTVDALDDGHLGCTIMPGFHLMKESLAVRTANLNRHAGKRYPPPTTIG	372
Qy	183	QATASGIAAAQAGAIVQWLAGRQR---YGAPEIYVAGGQWPEVQEAERLLAVTGAA	238
Db	373	NAVASGMDAVCGSV--MMHGRUKETGAGKPDVVIITGGG-----AAKVAEALPPA	423
Qy	239	FGATPQPTYLDSPVLDGLAAL-AAQG 263	
Db	424	FLAENTYRVADNLVIHGLNLIAAEG 449	
RESULT 8			
AAV38616			
ID	AAV38616	standard; Protein; 592 AA.	
XX	AAV38616;		
AC	AAV38616;		
XX			
XX	08-OCT-1999	(first entry)	
DE	Neisseria meningitidis	strain A antigen encoded by ORF61.	
XX			
DE	Neisseria meningitidis;	Neisseria gonorrhoeae; antigen; vaccine;	
KW	treatment; Neisseria	infection; meningitis; septicaemia; gonorrhea.	
KW			
XX	Neisseria meningitidis.		
OS			
XX			
XX	WO9924578-A2.		
PN			
XX			
PD	20-MAY-1999.		
XX			
PF	09-OCT-1998;	98WO-IB01665.	
XX			
PR	01-SEP-1998;	98GB-0019016.	
PR	06-NOV-1997;	97GB-0023516.	
PR	14-NOV-1997;	97GB-0024130.	
PR	18-NOV-1997;	97GB-0024386.	
PR	27-NOV-1997;	97GB-0025158.	
PR	10-DEC-1997;	97GB-0026147.	
PR	14-JAN-1998;	98GB-0000759.	
XX			
PA	(CHIR-) CHIRON SPA.		
XX			
PI	Grandi G, Masignani V, Pizzo M, Rappuoli R, Scarlato V;		
XX			
XX	WPI: 1999-327407/27.		
DR	N-PSDB; AAZ12074.		
DR			
XX	Proteins from Neisseria meningitidis and N. gonorrhoeae useful for		
PT	diagnosis, treatment and prevention of infection		
PT			
XX	Claim 4; Page 173; 524pp; English.		
PS			
XX			
CC	Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis		
CC	and N. gonorrhoeae antigenic proteins. They are encoded by open		
CC	reading frames (ORFs) AAZ11972-212358. The antigenic proteins,		
CC	their fragments, their nucleic acids and antibodies are used for		
CC	diagnosis, prevention (as vaccines) or treatment of Neisseria		
CC	infections, such as meningitis, septicaemia and gonorrhea. Both		
CC	organisms are closely related. Fragments of the nucleic acids		
CC	are useful as hybridisation probes and antisense reagents.		
XX			
SQ	Sequence	592 AA;	
Query Match		18.4%; Score 256.5; DB 20; Length 592;	
Best Local Similarity		32.0%; Pred.NO. 3.7e-18;	



Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for diagnosis, treatment and prevention of infection

Claim 4; Page 172; 524pp; English.

Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis* and *N. gonorrhoeae* antigenic proteins. They are encoded by open reading frames (ORFs) AA211972-212358. The antigenic proteins, their fragments, their nucleic acids and antibodies are used for diagnosis, prevention (as vaccines) or treatment of *Neisseria* infections, such as meningitis, septicæmia and gonorrhœa. Both organisms are closely related. Fragments of the nucleic acids are useful as hybridisation probes and antisense reagents.

CC	diagnoses, prevention (as vaccines) or treatment of <i>Neisseria</i>
CC	infections, such as meningitis, septicaemia and gonorrhea. Both
CC	organisms are closely related. Fragments of the nucleic acids
CC	are useful as hybridisation probes and antisense reagents.
XX	
SQ	Sequence 592 AA;
	Query Match 18.3%; Score 254.5; DB 20; Length 592;
	Best Local Similarity 31.6%; Pred. No. 6e-18;
	Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps:
QY	3 ILIDSGNSRLKVGWFDPAQOAAREPAVPAPVDNLDDLALGRWLATLPRRPOALGVNVAG 62
DB	341 LLLDGGNSRLUKWAVE--NGTATVGSAPYR---DLSPGLAEWAEKADGNVRIVGCavgc 393
QY	63 LARGEATAATL RAGGCDIRWLRAOPLAMGLRNGYRNPDLGADRWACWGVVLAROPSVHP 122
DB	396 EFKKAOVQEQAR---KIEWLPSSAQALGIRNHYRHPDEHSGSDRWFNALG--SRKRF--RN 449
QY	123 PLLVASGTAATLTDTIGFDNVFPGGILPGPAMRMGALAYGTAHPLADGLVADVPIDTH 182
DB	450 ACVVSCGTATVVDALDDGHYLGCTIMPGGHLMKESLAVRTANLNRIAGKKYPPPTTTG 509
QY	183 QAIASGTAAQAAGAIYQWLGRQR----YQOAPETIYVAGGQWPEVROEAERLLAVTGA 238
DB	510 NAVASGMDAVCGSVM--MHGRLUKEXTGACKPVDVITGG-----AAKVAELPPA 560
QY	239 FGATPQPTYLDSPVLDGLA--ALAAQG 263
DB	561 FLAENTVRYVADNLVIYGLLNNIAAEG 586

	RESULT 12
	AAV74912
ID	AAV74912 standard; Protein:592 AA:
QY	239 FGATPPTTILDSPLVDGLA-ALAAAG     :   :   :   :     :   :   :   :
Dd	561 FLAENTVRADNLVIYGLLNMIARAGEG

AAV/4912	AAV74912 standard	protein: 592 AA
ID	AAV74912 standard	protein: 592 AA

XX	
AC	AA74912;
XX	
DT	21-MAR-2000 (first entry)

XX	21-MAR-2000 (first entry)	Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
DT		
XX		
DE		
XX		
KW		Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW		antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW		antibacterial; gene therapy.

BE	<i>Neisseria meningitidis</i>	ORF	311	protein sequence	SEQ ID NO: 1298
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					
81					
82					
83					
84					
85					
86					
87					
88					
89					
90					
91					
92					
93					
94					
95					
96					
97					
98					
99					
100					

XX	<i>Neisseria meningitidis</i>
KW	<i>Neisseria meningitidis</i> ;
KW	antigenic; diagnosis; i
KW	antibacterial; gene the
XX	<i>Neisseria meningitidis</i> .
OS	

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy.

XX  
OS Neisseria meningitidis.

OS	Neisseria meningitidis
XX	
PN	WO9957280-A2

XX  
PN  
XX  
PD

11-NOV-1999.  
30-APR-1999.

30-APR-1999; 99WO-US09346

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 980S-0094869;  
PR 02-SEP-1998; 980S-0098994;  
PR 02-SEP-1998; 980S-0098994;  
PR 02-SEP-1998; 980S-0098994;

PR	02-SEP-1998;	980S-00990062.
PR	09-OCT-1998;	980S-0103749.
PR	09-OCT-1998;	980S-0103749.
PR	09-OCT-1998;	980S-0103749.

PR 09-OCT-1998; 98US-0103796.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.

PR	09-OCT-1998;	98US-0103796.
PR	25-FEB-1999;	99US-0121528.

XX



XX WO957280-A2.  
XX 11-NOV-1999.  
XX 30-APR-1999; 99WO-US09346.  
XX 01-MAY-1998; 98US-0083758.  
XX 31-JUL-1998; 98US-0094869.  
XX 02-SEP-1998; 98US-0098994.  
XX 09-SEP-1998; 98US-0099062.  
XX 09-OCT-1998; 98US-0103749.  
XX 09-OCT-1998; 98US-0103794.  
XX 09-OCT-1998; 98US-0103796.  
XX 25-FEB-1999; 99US-0121528.  
XX (CHIR ) CHIRON CORP.  
XX (GENO-) INST GENOMIC RES.  
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
XX Tettelin H, Venter JC;  
XX WPI: 2000-062150/05.  
XX N-PSDB; AA253671.  
XX Novel Neisserial polypeptides predicted to be useful antigens for  
XX vaccines and diagnostics  
XX Claim 2; Page 710; 1453pp; English.  
XX AA253015 to AA254336, AA254577 to AA254615, and AA274253 to AA275941  
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent  
XX PCR primers used in the exemplification of the present invention. The  
XX polypeptides, the polynucleotides, antibodies and compositions of  
XX the invention can be used as vaccines, as diagnostic reagents, and as  
XX immunogenic compositions. The polypeptides can be used in the  
XX manufacture of medicaments for treating or preventing infection due to  
XX Neisseria bacteria (e.g. meningitis and septicemia), to detect the  
XX presence of Neisseria bacteria, or to raise antibodies. They may also  
XX be used to screen for agonists or antagonists, which may themselves  
XX have use as antibacterial agents. The polynucleotides of the invention  
XX may also be used in gene therapy protocols.  
XX SQ Sequence 389 AA;  
Query Match 17.5%; Score 244; DB 21; Length 389;  
Best Local Similarity 31.5%; Pred. No. 4.3e-17;  
Matches 84; Conservative 44; Mismatches 113; Indels 26; Gaps 10;  
QY 3 ILDSGSRKLVGWFPDPAQAPAPAFVAFDNLDLALGRWLATLPRPQALGVNVAG 62  
DB LLLDGGNSRLKWAWE-NGTFATVGSAPYR-...DLSPGLAEWAERADGNVRIVGCACVG 192  
QY 63 LARGEIAATLRAGGCDIWLRAQPLAM-GLRNGYRNPDLGADRACWGVGLARQPSVH 121  
DB EFKKAQVQBLAR-...KIEWLPSSAQAALFGIRNHYRHPHSDRWFNALG--SRFS-R 246  
QY 122 PPLLVASFGTATTLDITGPNVPPGGLILPGPAMRGALAYGTAHLPLADGLVADYPIDT 181  
DB NACVVVCGGTAVTVADTDDGHTLYGGTIMPFGHLMRESLAVANLNHRHAKGYPPFTTT 306  
QY 182 HQAIAAGAAQAAGATVROWLAGRQR-...YGAPEIYVAGGCGWPEVQAEALLAVTGA 237  
DB 307 GNAVSGMMDVCGSVN--MMHGRLEKTKAGRPVDVITGGG-...AAKVAELPP 357  
QY 238 AFGATPQPTLYDSVLDGLA-ALAAQG 263  
DB AFLAENTVRVADNLVIYGLLNMAAEG 384

AAU91180  
ID AAU91180 standard; Protein; 242 AA.  
XX AC AAU91180;  
XX DT 05-JUN-2002 (first entry)  
XX DE Pantothenate kinase (Coax) #18.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Xylessa fastidiososa.  
XX PN WO200216601-A2.  
XX PD 28-FEB-2002...  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PI Yocum RR, Patterson TA;  
XX WPI: 2002-269358/31.  
XX N-PSDB; ABK54201.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
XX contacting composition comprising pantothenate kinase (Coax) protein  
XX with test compound and identifying inhibitor of the Coax protein -  
XX Claim 6; Page 111; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
XX comprising contacting an assay composition comprising a pantothenate  
XX kinase (Coax) protein with a test compound, and determining the ability  
XX of the test compound to inhibit the activity of the Coax protein, an  
XX essential enzyme for the production of coenzyme A. Coax protein is a  
XX valuable target for identifying bactericidal compounds. Coax modulating  
XX agents can be used in an infectious animal model to determine the  
XX efficacy, toxicity, or side effects of treatment with such an agent. This  
XX is the amino acid sequence of a pantothenate kinase (Coax) protein  
XX described in the invention.  
XX SQ Sequence 242 AA;  
Query Match 14.0%; Score 194.5; DB 23; Length 242;  
Best Local Similarity 27.9%; Pred. No. 3.9e-12;  
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;  
QY 4 LIDSGNSRLKVGWFPDPAQAPAREPA-...PVAFDNLDLALGRWLATLPRPQAL 56  
DB LFDLGNRFRKA-...SLREGVIGPVTVLPYLTETMDAFA-...LQELPRG-... 47  
QY 57 GVNVAAGLARGEA-...IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97  
DB 48 --RVAYLASVAAPAITTHVLELVKTHFEQVQVAATVA-...CAGVRIYA 92  
QY 98 NPQOLGADRACWGVGLARQPSVHPPLLVASFGTATTLDITGPDNVFPGGLLPGPAMMR 157  
DB HPERFGVDRETLALGSGY-...EGNVLVGVGTALTIDLLAANGCHLGRISASPTLMR 147  
QY 158 GALAYGTAHLPLADGLVADYPIDTHQAIASIAAQAQATVROWLAGRQVGAPEIYA 217  
DB 148 QALHARAEQLPLSGGNYLEFAEDTEDALVSCNGAAVALIERSYEAHQRLDQSVRLH 207  
QY 218 GGGWPEVQAEERLAVTGAAGFATPQPTLYDLSVLDGLAALAAOGA 264  
DB 203 GGG-...VASLLPWLG-...DVVHRPTL-...VLDGLAIWAAVA 240

Search completed: June 24, 2003, 21:46:23  
Job time : 29.4124 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.7386 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453a-15  
Perfect score: 1392  
Sequence: 1 MIIIDSGNSRLKVGWFDPD.....LDSPVLDGLAALAAQAGPTA 267

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	267	2 I40327	baf protein - Bord
2	256.5	18.4	592	2 H82031	probable biotin-la
3	254.5	18.3	592	2 B81009	Bira protein/Bvg a
4	194.5	14.0	242	2 A82637	conserved hypotet
5	178.5	12.8	248	2 H83111	hypothetical prote
6	161	11.6	259	2 AF1102	conserved hypotet
7	160.5	11.5	273	2 E97293	probable transcrip
8	155	11.1	259	2 AF1464	conserved hypotet
9	152	10.9	246	2 D72320	conserved hypotet
10	134.5	9.7	261	2 B87489	transcription acti
11	134	9.6	265	2 T36391	hypothetical prote
12	131.5	9.4	254	2 F83660	hypothetical prote
13	125	9.0	233	2 S66100	conserved hypotet
14	122	8.8	251	2 F83604	hypothetical prote
15	110.5	7.9	591	2 T44868	probable membrane
16	110.5	7.9	600	2 D87232	conserved membrane
17	109	7.8	636	2 T18542	moC protein - lep
18	107.5	7.7	480	2 F70785	hypothetical prote
19	107	7.7	4848	2 T30289	pristinamycin I sy
20	106.5	7.7	423	2 A36794	hypothetical prote
21	103.5	7.4	262	2 E75516	conserved hypotet
22	103	7.4	974	2 T35045	bacteriophage phiC
23	102.5	7.4	442	1 XUBRV5	3-phosphohikimate
24	102.5	7.4	465	2 H87301	conserved hypotet
25	102.5	7.4	603	2 A87342	conserved hypotet
26	101	7.3	309	2 B75565	hypothetical prote
27	101	7.3	566	2 T35203	probable two-compo
28	101	7.3	1110	1 S55279	guanylate cyclase
29	100.5	7.2	546	2 D83408	hypothetical prote

30 100.5 7.2 813 2 E87396  
31 100.5 7.2 821 2 AB0688  
32 100 7.2 308 2 F95903  
33 99.5 7.1 240 2 G84177  
34 99.5 7.1 274 2 H86937  
35 99.5 7.1 700 2 A84243  
36 99 7.1 280 2 B84219  
37 99 7.1 509 2 T34871  
38 99 7.1 974 2 T30204  
39 98.5 7.1 676 1 EDBE22  
40 98 7.0 439 2 I57561  
41 98 7.0 509 2 E85840  
42 97.5 7.0 7576 2 T17428  
43 97 7.0 257 2 S75559  
44 97 7.0 509 2 H90994  
45 97 7.0 562 2 H70688

## ALIGNMENTS

### RESULT 1

I40327  
baf protein - Bordetella pertussis  
C:Species: Bordetella pertussis  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 08-Oct-1999  
C:Accession: I40327; S70669  
R:DeShazer, D.; Wood, G.E.; Friedman, R.L.  
J. Bacteriol. 177, 3801-3807, 1995  
A:Title: Identification of a Bordetella pertussis regulatory factor required for tr  
A:Reference number: I40327; MUID:95325323; PMID:7601846  
A:Accession: I40327  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-267 <RES>  
A:Cross-references: EMBL:U12020; NID:9687228; PIDN:AAA75361.1; PID:9687229  
R:Allen, A.; Maskell, D.  
Mol. Microbiol. 19, 37-52, 1996  
A:Title: The identification, cloning and mutagenesis of a genetic locus required for  
A:Reference number: S70669; MUID:96419162; PMID:8821935  
A:Accession: S70669  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 239-267 <ALL>  
A:Cross-references: EMBL:X90711; NID:9992967; PIDN:CAA62242.1; PID:9992968  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
C:Genetics:  
A:Gene: baf

Query Match 100.0%; Score 1392; DB 2; Length 267;  
Best Local Similarity 100.0%; Pred. No. 7.2e-103;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIIIDSGNSRLKVGWFDPDQAARAPAPVAFDNLDLALGRLATLPPRORALGVNV 60  
Db 1 MIIIDSGNSRLKVGWFDPDQAARAPAPVAFDNLDLALGRLATLPPRORALGVNV 60  
QY 61 AGLARGEATAATLRAGGCCDIRWLRAQPLAMGLRNGYRNPDLGLADRWACWGVGLARQPSV 120  
Db 61 AGLARGEATAATLRAGGCCDIRWLRAQPLAMGLRNGYRNPDLGLADRWACWGVGLARQPSV 120  
QY 121 HPPLLVASFGTATTLDITIGPDNVFPGGLLPGPMRMGALAYTAHLPLADGLVADYPID 180  
Db 121 HPPLLVASFGTATTLDITIGPDNVFPGGLLPGPMRMGALAYTAHLPLADGLVADYPID 180  
QY 181 THQATASGTAAQAAGAIVRQWLAGRQRYCOAPEIYVAGGGWPPVROEAERLLAVTGAFA 240  
Db 181 THQATASGTAAQAAGAIVRQWLAGRQRYCOAPEIYVAGGGWPPVROEAERLLAVTGAFA 240  
QY 241 ATPQPTYLDSPVLDGLAALAAQAGPTA 267  
Db 241 ATPQPTYLDSPVLDGLAALAAQAGPTA 267

Query Match 18.3%; Score 254.5; DB 2; Length 592;  
Best Local Similarity 31.6%; Pred. NO. 1.5e-12;  
Matches 84; Conservative 44; Mismatches 113; Indels 25; Gaps 9;

Db 148 QALHARAEQLPLSGNLYFEAEDETDALVSCNGAANVALIERSYEAHORLDQSVRLDLH 207  
 QY 218 GGGWPEVQREARLAVTGAAGATPQPTTYLDSPLVDGLAALAAQGA 264  
 Db 208 GGG-----VASLLPWLG---DVVHREPL---VLDGLAIWAQAA 240

RESULT 5  
 H83111  
 hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83111  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 ;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 ;Reference number: A82950; PMID:20437337; PMID:10984043  
 ;Accession: H83111  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-248 <STO>  
 A:Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4279

Query Match 12.8%; Score 178.5; DB 2; Length 248;  
 Best Local Similarity 29.6%; Pred. No. 5.7e-07;  
 Matches 75; Conservative 32; Mismatches 93; Indels 53; Gaps 11;

QY 1 MIIIDSGNSRLKVGWFPDPAQAAREPAPVAFDNLDDALGRWLAT---LPRRPQALG 57  
 Db 1 MILEDCGSLIK--WRVIEG--AARSVAGGLAES--DDALVEQLTSQALPVRACRLVS 54  
 QY 58 VN-----VAGLARGAIAATLRAGCCDTRWLRQAQPLAGLNGYRNPDOLGADRWAC 109  
 Db 55 VRSEQETSQVLRLEQLFVPSLAVSSGKQL-----AGVRNGYLDYQRLGDLRWLA 105  
 QY 110 MVGV--LARQSVHPPLLVASFATTTDITGPDNVFPGGLILPGPAMRGALAYGTAHL 167  
 Db 106 LVAAHLAKK-----ACLVLDLGTAVTSDLVAAQGVHLGGYICPGNTLMRSQRLTHTRRI 160  
 QY 168 PLADGLVADYPIDTHQAIA--GTAAQA-----GAIVRWLAGRQRYGQAPAEI 214  
 Db 161 RYDDA-----EARRALASLPQGAATAEAVRGCLLMRLGRFVREQYAMACELLPDCEI 213  
 QY 215 YVAGGWPVEVQRE 227  
 Db 214 FLTGGDAELVRDE 226

RESULT 6  
 AF1102  
 conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e  
 C:Species: Listeria monocytogenes  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 C:Accession: AF1102  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
 Science 294, 849-852, 2001  
 A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
 ok, C.; Schlueter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A:Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; PMID:21537279; PMID:11679669  
 A:Accession: AF1102  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-259 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo0221  
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.6%; Score 161; DB 2; Length 259;  
 Best Local Similarity 26.3%; Pred. No. 1.5e-05;  
 Matches 78; Conservative 38; Mismatches 109; Indels 72; Gaps 16;

QY 1 MIIIDSGNSRLKVGWFPDPAQAAREPAPVAFDNLDDALGRWLATLPR-RPQRALGV- 58  
 Db 1 MILVIDGNTCTVGVIEQ-----KLLKHWRMTDRHRTSDELGMT 42  
 QY 59 -----NVAGLARGAIAATLRA--GGCDIRWLRQAQPLAM--GLRNGYR----- 97  
 Db 43 VLNFFSYANLTPSDIOGGIISSVPPIMHAMETMC--VRYFNIRPLVGVGKTKGLNLKVD 101  
 QY 98 NPDOLGADRWACMGVGLARQSPVHPPLLVASFATTTDITGPDNVFPGGLILPGPAMMR 157  
 Db 102 NPREIGSDR---IVNAVAASEYGPVIVVDFGTATTCYIDSGYVQGAIAAPGIMIST 158  
 QY 158 GALAYGTALPLADGLVADYPIDTHQAIA--SGIAAAQAG---AIVRW---LAGRQRYGQ 210  
 Db 159 EALYNRAAKLPRVD--IA-----ESSQIICKSVSSMQAGIFYGVGQCEGIIAEMKKQSN 212  
 QY 211 APEIYVAGGWPVEVQREARLAVTGAAGATPQPTTYLDS--PVLDDGLAALAAQCAPT 266  
 Db 213 ASPVVVATGGL--ARMITEKSSAV-----DILDFFLTLLKGLLELYRRNKPT 256

RESULT 7  
 E97293  
 probable transcription regulator, homolog of Bvg accessory factor [imported] - Clos  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: E97293  
 R:Nollings, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;  
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteriu  
 A:Reference number: A96900; PMID:21359325; PMID:21359325  
 A:Accession: E97293  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-273 <KUR>  
 A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC3200  
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.5%; Score 160.5; DB 2; Length 273;  
 Best Local Similarity 24.7%; Pred. No. 1.7e-05;  
 Matches 64; Conservative 42; Mismatches 104; Indels 49; Gaps 9;

QY 1 MIIIDSGNSRLKVGWFPDPAQAAREPAPVAFDNLDDALGRWLATLPRRQALGVN 59  
 Db 12 VILVLDVGNNTNVLGIYN-----DTKLTAEMRLSTDVLRSADEYGIQ 53  
 QY 60 VAGLARGAIAATLRAG-----GCCDIRWLRQAQPLAM--GLRNG-----YRN 98  
 Db 54 VNNLFQDDKLDPLTEGVVISSVVPNIMYSLEHMKRYKINFLVGVGKTKGINIKYDN 113  
 QY 99 PDOLGADRWACMGVGLARQSPVHPPLLVASFATTTDITGPDNVFPGGLILPGPAMMR 158  
 Db 114 PREVGADR---IVNAVAHEIKVRSIIIDFGTATTCFCAVRENGDYLGGAICPGIKVSSE 170  
 QY 159 ALAYGTALPLADGLVADYPI--DTHQAITASGIAAAQAGA---IVRWLAGRQRYGQAPE 213  
 Db 171 ALFERAAKLPVELIKPAYAICKNTISSIQSGIVYIGQVRYIVRMKBELOEKEKEP 230  
 QY 214 IYVAGGWPVEVQREARL 231

Db 231 LVTATGGLAKLISEAKNV 249

RESULT 8  
AF1464  
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Cl1p11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
Residues: 1-259 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16413682; GSPDB:GN00178  
A:Experimental source: strain Cl1p11262  
C:Genetics:  
A:Gene: lin0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.1%; Score 155; DB 2; Length 259;  
Best Local Similarity 25.9%; Pred. No. 4.3e-05;  
Matches 77; Conservative 38; Mismatches 110; Indels 72; Gaps 16;

QY 1 MILIDSGNSRLKGVHFDPAQAAREPAPAFDNLDLALGRLATLPR-RPQALGV- 58  
DB 1 MILVIDVGNCTGVYK-----EQLLRHWRMTDRHRTSDELGMT 42  
QY 59 -----NVAGLARGEIAATLRA--GGCDIRWLRAQPLAM--GLRNGYR--- 97  
DB 43 VLNFFSYANLTSDIOGIILSVVPPIMHAMETWC--VFENRPLVIGIKITGLNLKVD 101  
QY 98 NPDQGLADRWACMGVGLARQPSVHPPLVLVASFGTATLPTIGPDNVFPGGLILPGPAMMR 157  
DB 102 NPREIGSDR---IVNVAASEEYGTPIVVVDFGTATTCFYIDEAGYVQGAIPGMIST 158  
QY 158 GALAYCTAHLPLADGLVADYDPTDTHQAIASGIAAQAQ---AIVRQW---LAGRORYGQ 210  
DB 159 EALYNRAAKLPRVD--IA-----ESSQIIKSTVASMQAGIFYGFIQCGEIIAEMKQSN 212  
QY 211 APEIYVAGGWPVROEAERLLAVTGAAGATPQPTYLDLSPVLDGLAALAAQAGPT 266  
DB 213 TSPVVVATGGL--ARMITKSSAV-----DILDFLTAKGLELLYRNKPT 256

RESULT 9  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: D72320  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickley Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: D72320  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <ARN>  
A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g4981417  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0883  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 10.9%; Score 152; DB 2; Length 246;  
Best Local Similarity 25.0%; Pred. No. 7.1e-05;  
Matches 62; Conservative 36; Mismatches 88; Indels 62; Gaps 10;

QY 1 MILIDSGNS-----RLKVGWFPDPAQAAREPAPAFDNLDLALGRW 44  
DB 1 MYLLVDVGNTHSVFSTEDKTFRRWLSTGVFOTE-----DELFSH 42  
QY 45 LATLPRRQALGVNVNAGLARGEIAAATLRAAGGCDI-----RWLRAQPLAMGLRNG----- 95  
DB 43 LHPLLGDAMEEI-----KGIVASVVPQNTVIERFSQKYFHSPIWVKAKNGCVKW 94  
QY 96 -YRNPQGLADRWACMGVGLARQPSVHPPLVLVASFGTATLPTIGPDNVFPGGLILPGPA 154  
DB 95 NVKNPSEVGADR---VANVFAVKEYKNGIIDMTATTVDLV-VNGSYEGGAILPGFF 150  
QY 155 MMRGALAYCTAHLPLADGLVADYPI--DTHQAIASGIAAQAQAGAIVRQWLAGR--QRXGQ 210  
DB 151 MMVHSLFRGTAKLPLVEVKPADFVVGKDTENTRLGVVNGSVYAL--EGIIIGRIKEVYGD 208  
QY 211 APEIYVAG 218  
DB 209 LPVVLTTGG 216

RESULT 10  
B87489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87489  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87489  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CCI935  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 9.7%; Score 134.5; DB 2; Length 261;  
Best Local Similarity 28.8%; Pred. No. 0.0018;  
Matches 53; Conservative 18; Mismatches 78; Indels 35; Gaps 8;

QY 81 RWLRAQPLA-----MGLRNGYRNPQGLADRWACMGVGLARQPSVHPPLVLVASFGTATT 134  
DB 80 RYFNVPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAWVYPG---PLWVIDSGTATT 136  
QY 135 LDTIGPDNVFPGGLILPGPAMRGALAYGTAHPLADGLVADYDPTDTHQAIASGIAAQA 194  
DB 137 FDIVAADGAFEGGIIAPGINLSMQALHEAAKLP---RIAIQRPAGNRIVGTDTVSAMQS 193  
QY 195 GAIVRQWLAGRQRYGQAPETIYVAGGWPVROEAER-----LLAVTGAAGF---GATPQPT 246  
DB 194 GVF---W-----GYISLIEGLVA-----RIKAERCEPMTVIATGGVASFEGATDSID 238  
QY 247 YLDS 250  
DB 239 HFDS 242

RESULT 11  
T36391  
hypothetical protein SCE94.31c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 10.4407 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRLFTVDVGNSSVDIALWEG.....YDPLLVHGRMNLLYLHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	100.0	229	9	US-09-813-453A-12
2	1225	19.3	257	9	US-09-813-453A-53
3	193.5	16.6	246	9	US-09-813-453A-9
4	177.5	15.2	209	9	US-09-813-453A-21
5	167.5	14.3	262	9	US-09-813-453A-45
6	164.5	14.1	241	9	US-09-813-453A-63
7	158.5	13.6	258	9	US-09-813-453A-6
8	154.5	13.2	258	9	US-09-813-453A-2
9	148.5	12.7	258	9	US-09-813-453A-49
10	141.5	12.1	262	9	US-09-813-453A-11
11	139.5	11.9	255	9	US-09-813-453A-7
12	138.5	11.9	223	9	US-09-895-913A-74
13	138.5	11.9	223	9	US-09-813-453A-14
14	138.5	11.9	223	9	US-09-813-453A-67
15	138.5	11.9	254	9	US-09-813-453A-47
16	133.5	11.4	262	9	US-09-813-453A-8
17	133	11.4	256	9	US-09-813-453A-55
18	133	11.4	257	9	US-09-813-453A-13
19	132.5	11.3	244	9	US-09-813-453A-41

20	130	11.1	219	9	US-09-813-453A-57	Sequence 57, Appl
21	128.5	11.0	265	9	US-09-813-453A-4	Sequence 4, Appl
22	125.5	10.7	260	9	US-09-813-453A-51	Sequence 51, Appl
23	123.5	10.6	250	9	US-09-813-453A-3	Sequence 3, Appl
24	122.5	10.5	273	9	US-09-813-453A-10	Sequence 10, Appl
25	120.5	10.3	212	9	US-09-813-453A-59	Sequence 59, Appl
26	117	10.0	233	9	US-09-813-453A-17	Sequence 17, Appl
27	114	9.8	272	9	US-09-712-363-276	Sequence 276, App
28	114	9.8	272	9	US-09-813-453A-5	Sequence 5, Appl
29	108	9.2	592	9	US-09-813-453A-22	Sequence 22, Appl
30	103	8.8	592	9	US-09-813-453A-43	Sequence 43, Appl
31	99	8.5	460	9	US-09-813-453A-39	Sequence 39, Appl
32	98.5	8.4	248	9	US-09-813-453A-20	Sequence 20, Appl
33	95	8.1	242	9	US-09-813-453A-65	Sequence 65, Appl
34	92.5	7.9	249	9	US-09-813-453A-61	Sequence 61, Appl
35	89.5	7.7	249	9	US-09-813-453A-70	Sequence 70, Appl
36	87.5	7.5	595	9	US-10-072-036-69	Sequence 69, Appl
37	87	7.4	245	9	US-09-933-989A-7	Sequence 7, Appl
38	86	7.4	297	9	US-09-991-211-4	Sequence 4, Appl
39	85	7.3	461	9	US-10-142-835-35	Sequence 35, Appl
40	85	7.3	927	9	US-09-884-696-2	Sequence 2, Appl
41	84	7.2	274	10	US-09-815-242-10868	Sequence 10868, A
42	84	7.2	872	10	US-09-815-242-13594	Sequence 13594, A
43	84	7.2	976	10	US-09-801-368-248	Sequence 248, App
44	82	7.0	317	9	US-10-076-622-176	Sequence 176, App
45	82	7.0	317	9	US-09-551-621-176	Sequence 176, App

ALIGNMENTS

RESULT 1  
US-09-813-453A-12  
; Sequence 12, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Aquifex aeolicus  
US-09-813-453A-12

Query Match	100.0%	Score 1168;	DB 9;	Length 229;
Best Local Similarity	100.0%	Pred. No. 1.9e-107;		
Matches 229;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MRLFTVDVGNSSVDIALWEGKKVDFLKLSSHEEFKKEFPKLGALGIVSKQSFSEKVRGK	60	
DB	1	MRLFTVDVGNSSVDIALWEGKKVDFLKLSSHEEFKKEFPKLGALGIVSKQSFSEKVRGK	60	
QY	61	IPKIFLKKEFNFIQVDYKTPETLGTDRVALYSAKKFYGNVYVISAAGTALVIDLVLEG	120	
DB	61	IPKIFLKKEFNFIQVDYKTPETLGTDRVALYSAKKFYGNVYVISAAGTALVIDLVLEG	120	
QY	121	KFKGGFTTLGLGKKILSLDLAEGIPPEFFPEVIFLGRSTRCVGLGAYRESTFEIKST	180	
DB	121	KFKGGFTTLGLGKKILSLDLAEGIPPEFFPEVIFLGRSTRCVGLGAYRESTFEIKST	180	
QY	181	LKLWRKFKKFKVVIITGGEGKYSKFGIYDPLLVHGRMNLLYLHRI	229	
DB	181	LKLWRKFKKFKVVIITGGEGKYSKFGIYDPLLVHGRMNLLYLHRI	229	

Db 181 LKLRKVKFRKRVKITGEGKYSKFGIYDPLLVHRGMRNLLYLXRI 229

## RESULT 2

US-09-813-453A-53  
; Sequence 53, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 53  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Chlorobium tepidum  
US-09-813-453A-53

Query Match 19.3%; Score 225; DB 9; Length 257;  
Best Local Similarity 27.7%; Pred. No. 2.7e-14;  
Matches 71; Conservative 39; Mismatches 104; Indels 42; Gaps 8;  
QY 4 LTVDVGNSSVDIALWEGKVKDFLKLSHEEP-----LKEEF----- 39  
Db 3 LVVDIGNTSTLAIATGDEEPSVESPSALFADSDSTMREVFGNMARKHGEPQAIATCSVV 62  
QY 40 PKLKALGISVKOS-ESEKVRGKIPKIKFLKKNFPQVDYKTPETLGTDRVAL-AVSACK 97  
Db 63 PSATAVGSALLSLES-----VPVLTICCKLRFFRDLAYPHFTGADRLCALCAWSRHL 116  
QY 98 FYGKNVVISATALVIDIV-LEGKFKGGFTLGLGKKLILSDLAEGIPPEFFPERVEIF 156  
Db 117 FSEKPVIAVDIGTAITFDVLDVGNVYRGGLIMPGIDMAGALHSRTAQLPQVRIIDRPSL 176  
QY 157 LGRSTRECVLGAYRESFEFISTLKLVRKVKFRKF-----KVVITGGEGKY-----FSKF 207  
Db 177 LGRSTRECVLGAYRESFEFISTLKLVRKVKFRKF-----KVVITGGEGKY-----FSKF 207  
QY 208 GIYDPLLVHRGMRNLL 223  
237 SVIDEVLVRGSDLL 252

## RESULT 3

US-09-813-453A-9  
; Sequence 9, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 246  
; TYPE: PRT

; ORGANISM: Thermotoga maritima  
US-09-813-453A-9

Query Match 16.6%; Score 193.5; DB 9; Length 246;  
Best Local Similarity 26.5%; Pred. No. 3.3e-11;  
Matches 66; Conservative 45; Mismatches 105; Indels 33; Gaps 6;  
QY 3 FLTVDVGNSSVDIALWEGKVKDFLKLSHEEFLEE-----FPKLKALGI-- 47  
Db 2 YLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFOTEDLSHLHPLLDGDMREIKIGIVAS 61  
QY 48 -----SVKGSFSEKVRGKIPKIKFLKKNFPQVDYKTPETLGTDRVALAYSAKKFKYK 101  
Db 62 VVPTQNTVIERFSQKTFHISP--IWKAKGCVKNNVNFSEVGADRVANVAVFKEYGK 119  
QY 102 NVVVISATALVIDIVLEGKFKGGFTLGLGKKLILSDLAEGIPPEFFPERVEIFLGRST 161  
Db 120 NGIIDMGATATVYDLVNGSVYEGAILPGFFMVMVHSLFRGTAKLPLVEVKRPADVVGKDT 179  
QY 162 RECVLGGAYRESTEFITKSTLKLVRKVKFRKVVITGGEGKYS---KFIYDPLLVHRG 218  
Db 180 EENIRLGVNGSVYALEGIIGRIKEY-GDLPVVLTTGGQSKIVKDMIKHEIFDEDELTIGK 238  
QY 219 MRNLLYLX 227  
Db 239 -----VIH 241

## RESULT 4

US-09-813-453A-21  
; Sequence 21, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Campylobacter jejuni  
US-09-813-453A-21

Query Match 15.2%; Score 177.5; DB 9; Length 209;  
Best Local Similarity 25.8%; Pred. No. 1e-09;  
Matches 62; Conservative 44; Mismatches 83; Indels 51; Gaps 10;  
QY 4 LTVDVGNSSVDIALWEGK-----KVKDFLKLSHEEFLEEPPKALGISVKQSFSEKVRG 59  
Db 2 LUCDIGNSNANF-LDDNKYFTLNIDQFLFKNEQ-----KIFYVNEHLKHLK- 50  
QY 60 KIPKIKFLKKNFPQVDYKTPET-----LGTDRVALAYSAKKFKYKNNVVISATALVID 115  
Db 51 -----NQKNFINLEPYFLEDTIYQGLIDRIAACYTI-----EDGVVVDAGSAITID 97  
QY 116 LVLEGKFKGGFTLGLGKKLILSDLAEGIPPEFFPERVEI-FLGRSTRECVLGAYREST 174  
Db 98 IISNSIHGGFTLPGIANYKKIYSHISPRKSEFNTQVSLDAPQPTMDALSALGVFKGIY 157  
QY 175 EPIKSTLKLVRKVKFRKVVITGGEGKYSKFSF---GIYDPLLVHRGMR-----NLLY 224  
Db 158 LLIKDA-----AQNKLYFTGGDQFLANYFDHAIYDKLLIFRGMKKIENPNLLY 209



```
RESULT 5
US-09-813-453a-45
; Sequence 45, Application US/09813453a
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453a-45

Query Match      14.3%; Score 167.5; DB 9; Length 262;
Best Local Similarity 23.7%; Pred. No. 1.3e-08;
Matches 61; Conservative 50; Mismatches 103; Indels 43; Gaps 9;

QY 6 VDVGNSSVDIALWEGKVKDFLKLSEEFKEEPKLGKALGIVKQ-----SFSKVR 58
DB 5 LDVGNTAVLGVFEELGRHWRMETDRHKTED-----EYGMVYQLLEHGLSF-EDVK 58
QY 59 G-----KIPKIF-----LKENFPIQ-----VDYKTPETLGTDRVALAYS 96
DB 59 GIIVSSVPPIMFALERMCEKFKIKPLVVGPGIKTGINKIYENPREVGADRIYNAVAGI 118
QY 97 KFYGNVWVISAGTALVIDLVECK-FKGGFITLGLKGLKILSDLAEGIPPEFVEI 155
DB 119 HLYGSPILIVDFGATTTCYINEEKHYMGVITPGIMISAALYSRAAKLPRIETTPSS 178
QY 156 FLGRSTRECVLGGAYRESTEFIKSTLKLWRKVKRKFKVITGEGKYFSK----FGIYD 211
DB 179 VVGKNTVSAMQSGILYGVGVGVEGIVKRMKEAKQEPKVIATGGLAKLISEESNVIDVD 238
QY 212 PLLVHRGMNLLYLYHR 228
DB 239 PFLTLAG---LYMLYR 252

RESULT 6
US-09-813-453a-63
; Sequence 63, Application US/09813453a
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; PRIOR FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Thioacillus ferrooxidans
US-09-813-453a-63
```

```
Query Match      14.1%; Score 164.5; DB 9; Length 241;
Best Local Similarity 29.5%; Pred. No. 2.3e-08;
Matches 46; Conservative 31; Mismatches 70; Indels 9; Gaps 4;

QY 77 DYKTPETLGTD-RVALAYSARKFYGNVWVISAGTALVIDLVEGKFKGGFTTLGLGKKL 135
DB 90 DYHPESLGFDRCCLLAAMDYPQDSIVIDMGTAITIDLLAGGHFRGRILPGIAMS 149
QY 136 KILSDLAEGIPPEFPEVEIFLGRSTRECVLGGAYRESTEFIKSTLKLWRKVKRKFKV 195
DB 150 RGLHEGTALLPEVLNAPAEMLGNDTSNAIQAGVHLFADALRGAITDPRQ-YSPQARIL 208
QY 196 ITGGEKGYFS---KFGIYDPLLVHGRMNLLYLYHR 228
DB 209 ITGGAERMQPFIAGSLYOPHLLRG----FYLWIR 240

RESULT 7
US-09-813-453a-6
; Sequence 6, Application US/09813453a
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813.453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR FILING DATE: 2000-08-24
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-09-813-453a-6

Query Match      13.6%; Score 158.5; DB 9; Length 258;
Best Local Similarity 22.0%; Pred. No. 1e-07;
Matches 57; Conservative 48; Mismatches 113; Indels 41; Gaps 9;

QY 4 LTVDVGNSSVDIALWEGKVKDF-----LKLSEEFKKEEF-----PKLKALGIVKQS 52
DB 3 LCIDGNTNTVFSVWDG---TDFATWRIATDHRRTADEYFVWLNTLMQLKGLQGRISA 59
QY 53 FSEKVRGKIPKIKFLKKE-----NFFIQVDYKTPETLGTDRVALAYS 94
DB 60 I---ISSAPRVVFNLRVLCNRYFDCRPVYVVGKGCLEVPAPRVDPGTTPORLVNTVA 116
QY 95 AKKFGKNVWVISAGTALVIDLVL-EGKFGKGFITLGLKGLKILSDLAEGIPPEFPEV 153
DB 117 GYDRGGDLIVDFGTATTFDVVAPDGAYIGVAPGVNLSLEALHMAAALPHVDVTKP 176
QY 154 EIFLGRSTRECVLGGAYRESTEFIKSTLKLWRKVKRKFKVITGEGKYFS-KFGIYDP 212
DB 177 QGVIGTNTVACIOSGVWYIGLVGVEIVQIRMERDRPMKVIATGGLASLFDLGLDFDK 236
QY 213 L---LVHGRMNLLYLYHR 228
DB 237 VEDDLTMHGLR-LIFDYNK 254

RESULT 8
US-09-813-453a-2
; Sequence 2, Application US/09813453a
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 6.6898 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-12

Perfect score: 1168

Sequence: 1 MRELTVDGNSVDIALWEG.....YDPLLVRGMRNLLLYHRI 229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

#### Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.5	7.4	351	4	US-09-457-040B-6
2	86	7.4	297	4	US-09-632-947B-4
3	85.5	7.3	350	4	US-09-457-040B-37
4	84.5	7.2	872	2	US-08-844-057-2
5	84.5	7.2	872	4	US-09-006-730-2
6	81.5	7.0	311	4	US-09-175-172-2
7	81.5	7.0	1019	4	US-09-434-066-23
8	81	6.9	533	4	US-09-603-185-2
9	79	6.8	321	2	US-08-922-146-2
10	79	6.8	321	2	US-08-922-146-4
11	79	6.8	321	4	US-09-343-986-2
12	79	6.8	321	4	US-09-343-986-4
13	79	6.8	321	4	US-09-322-081-2
14	79	6.8	321	4	US-09-322-081-4
15	79	6.8	340	1	US-08-097-831-2
16	79	6.8	392	4	US-09-026-408-2
17	79	6.8	405	4	US-08-026-408-13
18	79	6.8	406	1	US-08-434-881-2
19	79	6.8	406	3	US-08-977-771-2
20	79	6.8	406	4	US-09-361-773-2
21	78.5	6.7	247	4	US-08-858-207A-300
22	78	6.7	252	4	US-09-071-035-20
23	78	6.7	272	4	US-09-071-035-18
24	78	6.7	324	1	US-08-597-236-10
25	78	6.7	324	1	US-08-746-682A-10
26	77.5	6.6	721	4	US-09-134-078-19
27	77	6.6	414	2	US-08-599-171A-25

Sequence 25, Appl  
Sequence 25, Appl  
Sequence 25, Appl  
Sequence 29, Appl  
Sequence 9, Appl  
Sequence 42, Appl  
Sequence 47, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 4575, Ap  
Sequence 3701, Ap  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 6, Appl  
Sequence 21, Appl  
Sequence 8, Appl  
Sequence 4, Appl

28 77 6.6 414 2 US-08-646-590B-25  
29 77 6.6 414 3 US-09-069-226-25  
30 77 6.6 414 4 US-09-412-184-25  
31 76 6.5 537 3 US-09-028-934-29  
32 75.5 6.5 459 4 US-09-129-112-9  
33 75 6.4 511 4 US-08-676-444-42  
34 75 6.4 857 4 US-08-887-534A-47  
35 74 6.3 277 1 US-08-400-413-1  
36 74 6.3 911 4 US-08-461-562B-2  
37 73.5 6.3 281 4 US-09-134-001C-4575  
38 73.5 6.3 383 4 US-09-134-001C-3701  
39 73 6.2 776 1 US-08-021-601-2  
40 73 6.2 776 1 US-08-082-849B-2  
41 73 6.2 776 5 PCT-US94-01624-2  
42 73 6.2 1022 1 US-08-271-364A-8  
43 73 6.2 1022 2 US-08-222-715B-27  
44 72.5 6.2 509 3 US-08-822-324-8  
45 72.5 6.2 1009 4 US-09-693-146-4

#### ALIGNMENTS

##### RESULT 1

US-09-457-040B-6  
; Sequence 5, Application US/09457040B  
; Patent No. 6387641

GENERAL INFORMATION:  
; APPLICANT: Vertex Pharmaceuticals Incorporated  
; APPLICANT: Bellon, Steve

; TITLE OF INVENTION: Crystallized P38 Complexes

; FILE REFERENCE: VPI/98-14

; CURRENT APPLICATION NUMBER: US/09/457,040B

; CURRENT FILING DATE: 1999-12-08

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Mouse

US-09-457-040B-6

Query Match 7.4%; Score 86.5; DB 4; Length 351;

Best Local Similarity 18.7%; Pred No. 0.6;

Matches 53; Conservative 46; Mismatches 74; Indels 111; Gaps 13;

QY 8 VGNSSVDIALWEGKKVKQDFKLKLSHEEFLK-----EEFKLKALG----- 46  
Db 1 MGNAAAARKGSEQSVKEFLAKAKEDFLKKWETPSQNTAQDQDFRIKTLGTGSGFGRVML 60  
QY 47 ISVKQS---FSEKV--RGKIPKIK-----FLKKENFP-----IQVDYKTPETLGT 87  
Db 61 VKHKEGSGNHAKILDKQKVKLKQLEHTLNKRILOAVNFPFLVKLEFSFKNSNL--- 117  
QY 83 RVALAYSA-----KKEYG-----KNVVVIS 107  
Db 118 YVMVEYVAGGEMFSLHRLRGRSEPHARYAAQIVLTFEYHLSLDLIYRDLKPENLIDQ 177  
QY 108 AGTALVIDLVLEKFGKGGFITL-----GLGKKLK-----ILSLAEGIP 146  
Db 178 QGYIQVDFGFAKRYKVGRTWTLGGTPEYLAPEILSKGYNKAVDWAALGVLIYEMAAGYP 237  
QY 147 EFPPEE-VEIFIGRSTRECVLGGAYNRETFEIKSTLUKLRKVPK 189  
Db 238 PFFADQPIQIY-----EKIVSGKRVFPFSPHSSDLKDLRLNLIQ 275

##### RESULT 2

US-09-632-947B-4

; Sequence 4, Application US/09632947B

; Patent No. 6356845

GENERAL INFORMATION:

; APPLICANT: Pharmacia & Upjohn

```

148 QY      148 FPFE-VEIFLGRSTCVLGGAAYRESTETFKSTLKLRKVKFK 189
    || : :| | : :| | : :| | : :| | : :|
238 DB      238 FFAQDPIQIY-----EKIVSGKYRFPFSHFSSDLKDILRLNLIQ 274

RESULT 4
US-08-844-057-2
; Sequence 2, Application US/08844057
; Patent No. 5863777
; GENERAL INFORMATION:
;   APPLICANT: Lawlor, Elizabeth
;   TITLE OF INVENTION: No. 5863777el Compounds
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Smithkline Beecham Corporation
;     STREET: 709 Swedeland Road
;     CITY: King of Prussia
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19406-0939
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/844,057
;     FILING DATE: 18-APR-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 9607991.8
;     FILING DATE: 18-APR-1996

```

```

NAME: GUMMAL, EDWARD R
REGISTRATION NUMBER: 38,991
REFERENCE/DOCKET NUMBER: P31458-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-057-2

Query Match          7.2%; Score 84.5; DB 2; Length 872;
Best Local Similarity 27.7%; Pred. No. 3.3;
Matches 46; Conservative 25; Mismatches 52; Indels 43; Gaps 10;

QY      29 LSHEFLKEEFFPKLKAIGISVKVSFEKVGKPIKIFLKKNFPQIVDYKTPTLTGDR 88
        |: |::| ::||:: ||::| :||::| ::||::| :||::| :||::|
DDB     591 LNEVEFLRFDFTHFOAVTAELRAIEQQNKKIWEALEV-----TVETDDITAKEMG--- 643

QY      89 VALAYSARKFYKNNVVVISAGTALVIDLVLE---GRFKGGFITGLGCKKLKILSDLAEGI 145
        |: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DDB     644 -AIALFGEK-YGVARVRVTIG----DYSIELCGGTHVGNWTSSELGLFKIVK-----EEGI 691

QY      146 PEFFPEEVEIFIGRSTRE--CVLG-----GAYRESTEFIK---STLK 182
Db       692 -----GSQTFRRLIATVTKGAFAFYRROEDALKAVAATLK 725

```

RESULT 5  
US-09-006-730-2  
; Sequence 2, Application US/09006730  
; Patent No. 6410285  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 6410285el  
; Compounds

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Smithline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/006,730  
FILING DATE:  
CLASSIFICATION: 435  
Prior Application Data:  
APPLICATION NUMBER: US/08/844,057  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: 9607991.8  
FILING DATE: 18-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-006-730-2

Query Match 7.2%; Score 84.5; DB 4; Length 872;  
Best Local Similarity 27.7%; Pred. No. 3.3;  
Matches 46; Conservative 25; Mismatches 52; Indels 43; Gaps 10;

Qy 29 LSHEEFLKEEFPKLGKALGKISVQSEKVRGKIPKIKLKKENFPQVDYKTPETLGTDR 88  
Db 591 LNEVEFLRFDTHFOAVTAELRAEQVNEKINEALEVK---TVETDIDTAKEMG--- 643  
Qy 89 VALAYSARKFYKGNVVISAGTALVDLVLE---GKFKGGFTTLGLGKLLKILSDLAEGI 145  
Db 644 -AIALFGKEK-YGKAVRVVTIG-----DYSIELCGGTHVNTSEIGLFLKIVK-----EEGI 691  
Qy 146 PEFPPEVEIFLGRSTRE--CVLG-----GAYRESTEFIK---STLK 182  
Db 692 -----GSTRILAVTGKEAFEAYREQEDALKAVAATLK 725

RESULT 6  
US-09-175-172-2  
; Sequence 2, Application US/09175172  
; Patent No. 6255075  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Jianzhong  
; APPLICANT: McDevitt, Damien  
; TITLE OF INVENTION: bira  
; FILE REFERENCE: GM10172  
; CURRENT APPLICATION NUMBER: US/09/175,172  
; CURRENT FILING DATE: 1998-10-20  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 311  
; TYPE: PPT  
; ORGANISM: Streptococcus pneumoniae

US-09-175-172-2

Query Match 7.0%; Score 81.5; DB 4; Length 311;  
Best Local Similarity 18.9%; Pred. No. 1.7;  
Matches 54; Conservative 40; Mismatches 67; Indels 125; Gaps 14;

Qy 20 GKVKVDFLKLSSHEEFLKEEFPKLGKALGI---SVKQSFSEKVRGK--IPKIKLKKENFPPI 74  
Db 20 GEKIAEKLSLRTSIWK-AIKRLEQEGIEIDSINKRGYKLMNGDLILPEI---LEENLPI 75  
Qy 75 QVDYTPETLGT-----DRVALAYS-----KKFVG----- 100  
Db 76 KVSFK-PETKSTQLDKAKEAIDLGHEANTLYLASYQTAGRCRFORFSYSPQGGIYMTLHLK 134  
Qy 101 -----KNVVVISAGTALVDLVLEKFKGFTTLGL-GKLLK 136  
Db 135 PNLPYDKLPSYTLVAGAVYKAIKNTLIDVDIKWYNDIYLNHNKIGGILTEAMTSVETG 194  
Qy 137 ILSDLAEGIP-----EFPPEVEIFLGRSTRECVLGAGYRESTEFIKST-----LK 182  
Db 195 LVTDIIGVGINFITKIDFPQELK-----EKAASLFKATAPITRNEILIE 238  
Qy 183 LWRKVFRRKFKVVITGGECKYFSKFGIYDPLLVHGRMRLLYLHR 228  
Db 239 IWRATFE-----TPAE-----ELLYLYKK 257

RESULT 7  
US-09-434-066-23  
; Sequence 23, Application US/09434066  
; Patent No. 6465714  
; GENERAL INFORMATION:  
; APPLICANT: Luthman, L. Holger  
; APPLICANT: Galli, L.G. Joakim  
; TITLE OF INVENTION: Congenic Animal Models of No. 6465714-Insulin  
; TITLE OF INVENTION: Dependent Diabetes Mellitus  
; FILE REFERENCE: 09705/009001  
; CURRENT APPLICATION NUMBER: US/09/434,066  
; CURRENT FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 1019  
; TYPE: PPT  
; ORGANISM: Rat  
US-09-434-066-23

Query Match 7.0%; Score 81.5; DB 4; Length 1019;  
Best Local Similarity 23.4%; Pred. No. 8.3;  
Matches 54; Conservative 31; Mismatches 69; Indels 77; Gaps 14;

Qy 32 BEFLKREEFKLGKALGISVK-----QSFSEKVRGK-----IPKIKLKKENFPPIQVDY 78  
Db 407 QEMVFECKDLNAVAFRPFKDKERPRGYTSKIAGKLHYVPLNGVLTAEYLLLEERPDLD- 465  
Qy 79 KTPETLGTDRVALAYSARKFYKGNVVIS-AGT-----ALVIDLV-----LEGKFKGG 125  
Db 465 MVLDKLRPENVRVAIVSKSFEGKTRTEQMYGTQYKQEAIPEDVIQKWNADLNGKFK-- 523  
Qy 126 FITLGLGKLLKILSDLAEGIP---EFPPEVEIFLGRSTRECVLGAGYRESTEF---IKS 179  
Db 524 -----LPTKNEFIPNTNFEL-----ALEKDATYPALIKD 553  
Qy 180 TL--KLWRKVFRRKFKVVITGG--EGKYFSKFGIYDPLLVHGRMRLLYLY 226  
Db 554 TMSKLM---FKQDDKFFLPKACLNFEPFSPFAYVDPLHC-----NWAILY 596

RESULT 8  
US-09-603-185-2  
; Sequence 2, Application US/09603185  
; Patent No. 6271004  
; GENERAL INFORMATION:





TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-922-146-4

Query Match 6.8%; Score 79; DB 2; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVDFLKLKLSHEEFLKEEFPKALKALGISVKQSFSEKVRG 59  
DB 72 IDLGNNITIKDVEAAIKMAHNTIQHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKENFPIQVDYKTPETLGTDRVALAYSAGK---KFGKNVWV-----ISAGTAL 112  
DB 122 QRIYVNAKNNDLVFGIG---PAGTGKTFLLAVVYAAKQLRGAGKRIYLTTPRAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFTLGLGKKLILSDLAEGIPPEFPPEVEIFLGRSTRE 163  
b 179 GFPLGDLKEKVDYPLRPDYGLTYVLGREQTERFIE---RGIETIAP--LAYMRGRTLED 233  
QY 164 CVLGAGYRESTEFIKSTLKLWRKFKRVKVVITGGEGKYSKFGIYDPLLVHGRMNL 223  
DB 234 AFVILDEAQNTHAQMKNFLTRLGFGS--KMWVTGDQTO-----IDLPGKVGKSLKEAV 285  
QY 224 YLYHRI 229  
DB 286 SRLHN 291

RESULT 11  
US-09-343-986-2  
Sequence 2, Application US/09343986  
Patent No. 6248863  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K.R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: phoH homolog  
FILE REFERENCE: P50444-3-D2  
CURRENT APPLICATION NUMBER: US/09/343,986  
CURRENT FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 08/922,146  
PRIOR FILING DATE: 1997-08-25  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 321  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-343-986-2

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVDFLKLKLSHEEFLKEEFPKALKALGISVKQSFSEKVRG 59  
DB 72 IDLGNNITIKDVEAAIKMAHNTIQHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKENFPIQVDYKTPETLGTDRVALAYSAGK---KFGKNVWV-----ISAGTAL 112  
DB 122 QRIYVNAKNNDLVFGIG---PAGTGKTFLLAVVYAAKQLRGAGKRIYLTTPRAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFTLGLGKKLILSDLAEGIPPEFPPEVEIFLGRSTRE 163  
DB 179 GFPLGDLKEKVDYPLRPDYGLTYVLGREQTERFIE---RGIETIAP--LAYMRGRTLED 233  
QY 164 CVLGAGYRESTEFIKSTLKLWRKFKRVKVVITGGEGKYSKFGIYDPLLVHGRMNL 223  
DB 234 AFVILDEAQNTHAQMKNFLTRLGFGS--KMWVTGDQTO-----IDLPGKVGKSLKEAV 285  
QY 224 YLYHRI 229

Db 286 SRLHN 291

RESULT 12  
US-09-343-986-4  
Sequence 4, Application US/09343986  
Patent No. 6248863  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K.R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: phoH homolog  
FILE REFERENCE: P50444-3-D2  
CURRENT APPLICATION NUMBER: US/09/343,986  
CURRENT FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 08/922,146  
PRIOR FILING DATE: 1997-08-25  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 321  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-343-986-4

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVDFLKLKLSHEEFLKEEFPKALKALGISVKQSFSEKVRG 59  
DB 72 IDLGNNITIKDVEAAIKMAHNTIQHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKENFPIQVDYKTPETLGTDRVALAYSAGK---KFGKNVWV-----ISAGTAL 112  
DB 122 QRIYVNAKNNDLVFGIG---PAGTGKTFLLAVVYAAKQLRGAGKRIYLTTPRAVEAGESL 178  
QY 113 -----VIDLVLEGKFGGFTLGLGKKLILSDLAEGIPPEFPPEVEIFLGRSTRE 163  
DB 179 GFPLGDLKEKVDYPLRPDYGLTYVLGREQTERFIE---RGIETIAP--LAYMRGRTLED 233  
QY 164 CVLGAGYRESTEFIKSTLKLWRKFKRVKVVITGGEGKYSKFGIYDPLLVHGRMNL 223  
DB 234 AFVILDEAQNTHAQMKNFLTRLGFGS--KMWVTGDQTO-----IDLPGKVGKSLKEAV 285  
QY 224 YLYHRI 229  
DB 286 SRLHN 291

RESULT 13  
US-09-322-081-2  
Sequence 2, Application US/09322081  
Patent No. 6323336  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: No. 6323336el phoH Homolog  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch St  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSEQ for Windows Version 2.0b

;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/322,081  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/922,146  
;; FILING DATE: 25-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson, Arthur E  
;; REGISTRATION NUMBER: 34,354  
;; REFERENCE/DOCKET NUMBER: P50444-03  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609/520-3254  
;; TELEFAX: 609/520-3259  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 321 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-322-081-2

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVDFLKLKLSHEEFLKEEPPKLGKALGKISVKQSFSEKVRG 59  
DB 72 IDLGNNTIKDVEAAIKMAHNTTIOHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKENPFIQVDYKTPETLGTDRVALAYSAGK---KFGKKNVVV-----ISAGTAL 112  
DB 122 QRIYVNAKNNDLVFGIG---PAGTKTFLAVVYAAKQLRGKAGKRIVLTRPAVEAGESL 178  
QY 113 -----VIDLVLEGKFKGGFTTLGLGKKILSDLAEGIPPEEVEIFLGRSTRE 163  
DB 179 GFLPGDLKEKVDPLRYDGLTYVLGREGTERFIE---RGIIETAP--LAYMRGRTLED 233  
QY 164 CVLGGAYRSTEFIKSTLKLWRKVKFKKVVITGGEGKFKYFSGFYDPLLVHVRGMNLL 223  
DB 234 AFVLDEAQNTHAQMKNMFLTRLGFGS--KMYVTGDQTO-----IDLPGKVKSGLKEAV 285  
QY 224 YLVHRI 229  
DB 286 SRLHNV 291

RESULT 14  
US-09-322-081-4  
Sequence 4, Application US/09322081  
Patent No. 6323336  
GENERAL INFORMATION:  
APPLICANT: Burnham, Martin K. R.  
APPLICANT: Lonetto, Michael A.  
APPLICANT: Warren, Patrick V.  
TITLE OF INVENTION: No. 6323336el phoH Homolog  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/322,081  
FILING DATE:

;;  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/922,146  
;; FILING DATE: 25-AUG-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jackson, Arthur E  
;; REGISTRATION NUMBER: 34,354  
;; REFERENCE/DOCKET NUMBER: P50444-03  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 609/520-3254  
;; TELEFAX: 609/520-3259  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 321 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-09-322-081-4

Query Match 6.8%; Score 79; DB 4; Length 321;  
Best Local Similarity 20.7%; Pred. No. 3.1;  
Matches 51; Conservative 42; Mismatches 105; Indels 48; Gaps 10;  
QY 6 VDVGNS-----SVDIALWEGKKVDFLKLKLSHEEFLKEEPPKLGKALGKISVKQSFSEKVRG 59  
DB 72 IDLGNNTIKDVEAAIKMAHNTTIOHLLDLYDEEITKDAFGK-----TIRAKTMG 121  
QY 60 KIPKIKFLKENPFIQVDYKTPETLGTDRVALAYSAGK---KFGKKNVVV-----ISAGTAL 112  
DB 122 QRIYVNAKNNDLVFGIG---PAGTKTFLAVVYAAKQLRGKAGKRIVLTRPAVEAGESL 178  
QY 113 -----VIDLVLEGKFKGGFTTLGLGKKILSDLAEGIPPEEVEIFLGRSTRE 163  
DB 179 GFLPGDLKEKVDPLRYDGLTYVLGREGTERFIE---RGIIETAP--LAYMRGRTLED 233  
QY 164 CVLGGAYRSTEFIKSTLKLWRKVKFKKVVITGGEGKFKYFSGFYDPLLVHVRGMNLL 223  
DB 234 AFVLDEAQNTHAQMKNMFLTRLGFGS--KMYVTGDQTO-----IDLPGKVKSGLKEAV 285  
QY 224 YLVHRI 229  
DB 286 SRLHNV 291

RESULT 15  
US-08-097-831-2  
Sequence 2, Application US/08097831  
Patent No. 5510473  
GENERAL INFORMATION:  
APPLICANT: Camerini-Otero, Rafael D.  
APPLICANT: Angov, Evangelina  
TITLE OF INVENTION: Cloning and Expression of Tag reca  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KNOBE, MARTENS, OLSON AND BEAR  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,831  
FILING DATE: 19930726  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115

```
; REFERENCE/DOCKET NUMBER: NIH066.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-097-831-2

Query Match          6.8%; Score 79; DB 1; Length 340;
Best Local Similarity 24.2%; Pred. No. 3.4;
Matches 40; Conservative 25; Mismatches 50; Indels 50; Gaps 7;

Oy 66 FLKENPPIQVDYKTPETLGTDRVALAYSARKFYGKNVVISAGTALVIDLVLECK-FKG 124
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
-Db 189 FINQVREKVGVMYGNPETTPGGR-----ALKFY-----SSVRLDVKSGQPIKV 232

   / 125 GFITLGLGKKLILSD-LAEGIPEFFPEEVEIFLGR----- 159
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 233 GNEAVGIKVKVYKVKLA---PPFREAELEIYFGRGLDPYMDLVNVAAGVIEKAGSW 289

Oy 160 -STRECVLGGAYRESTEFIKSTLKLWR---KVEKRKKVITGS 199
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 290 FSYGHRIGOGCKEAKAETLRPERPELLLEIRAKVLERADKVVVLAAG 334
```

Search completed: June 24, 2003, 21:49:26  
Job time : 7.6898 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 18.6773 seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRFLTVGVGSSVDIALWEG.....YDPLLVRGMRNLLIYHRI 229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	100.0	229	16	O67753 aquifex aeo
2	193.5	16.6	246	16	Q9WZY5 thermotoga
3	177.5	15.2	209	16	Q9PIA9 campylobact
4	156.5	13.4	259	16	Q8YAC5 listeria mo
5	149.5	12.8	259	16	Q92F54 listeria in
6	145	12.4	256	16	Q8RFE4 fusobacteri
7	141.5	12.1	262	16	O51477 borrelia bu
8	138.5	11.9	224	16	O25533 helicobacte
9	138.5	11.9	254	16	Q9KGH5 bacillus ha
10	138.5	11.9	273	16	Q97EB4 clostridium
11	137.5	11.8	258	2	Q9F985
12	136.5	11.7	255	16	Q8R7M2 thermotoga
13	133.5	11.4	262	16	Q9RX54 delinococcus
14	133	11.4	257	16	P74045 synecocyst
15	131.5	11.3	223	16	Q9ZKY6 helicobacte
16	129.5	11.1	261	16	Q9A6Z1 caulobacter

```

17 128.5 11.0 265 16 Q9XBN6
18 122.5 10.5 273 16 O83446
19 120.5 10.3 212 2 O32514
20 120.5 10.3 259 16 O8XHL5
21 117 10.0 276 16 O8YQD7
22 114 9.8 272 16 O06282
23 113.5 9.7 295 16 O8Y2M4
24 112 9.6 274 16 Q9CD56
25 108 9.2 592 16 Q9JW17
26 105 9.0 224 16 Q98Q93
27 103 8.8 592 16 Q9JXF1
28 102.5 8.8 254 17 O28311
29 101.5 8.7 476 16 O8RH37
30 100.5 8.6 763 16 O8XJY4
31 98.5 8.4 248 16 Q9HWC1
32 97.5 8.3 484 16 Q9ZMF7
33 96.5 8.3 328 10 Q9XQB0
34 96 8.2 543 5 Q9U4D6
35 96 8.2 578 16 Q97H45
36 95.5 8.2 426 17 O57987
37 95 8.1 242 16 Q9PCI4
38 95 8.1 390 5 O8T8V5
39 95 8.1 816 16 Q8RB00
40 94.5 8.1 484 16 O25055
41 94 8.0 483 8 Q9GER7
42 93.5 8.0 427 3 Q9URY5
43 93.5 8.0 613 16 Q8RDP8
44 93 8.0 267 12 Q8V9A4
45 93 8.0 487 8 Q9GEQ9

```

#### ALIGNMENTS

#### RESULT 1

```

O67753 PRELIMINARY; PRT; 229 AA.
ID AC O67753;
AC O67753;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein AQ_1924.
GN AQ_1924.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
RX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5; PubMed=9537320;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.";
RL Nature 392:353-358(1998).
DR EMBL; AE000763; AAC07720.1;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvq_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; complete proteome.
SQ SEQUENCE 229 AA; 26068 MW; 64C347A285FABE9A CRC64;

```

```

Query Match 100.0%; Score 1168; DB 16; Length 229;
Best Local Similarity 100.0%; Pred. No. 5.1e-87;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRFLTVGVGSSVDIALWEGKKYKDFLKLSEEFLEKFPKPKALGIVKQSFSEKVRGK 60
    |||||
Db 1 MRFLTVGVGSSVDIALWEGKKYKDFLKLSEEFLEKFPKPKALGIVKQSFSEKVRGK 60
    |||||
QY 61 IPKIKFLKENFPLOVDYKTPETLGTDRVALAYSAKKFKYKKNVVVISAGTALVIDLVLEG 120
    |||||

```



```
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL591974; CAD00748.1; -
DR MEROPS; M41.009; -
DR ListiList; LMO00221; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F1068049D80025B7 CRC64;

Query Match 13.4%; Score 156.5; DB 16; Length 259;
Best Local Similarity 23.8%; Pred. No. 6.4e-05;
Matches 61; Conservative 39; Mismatches 113; Indels 43; Gaps 7;

QY 4 LTVDVGNSSVDIALWEKKVKDFLKLKLSHEEFKKEFPKALKALGISVKQSFSEK----- 56
jB 1:||||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 LVIVIDGNTNCTGVYKEQKL-----LKHWRMTDRHRTSDGLGTMVLFNFFSYANLTPSDI 57
QY 57 ----VRGKIPKIKELK-----KENFPIDVDYKTPETLGTDRVALAY 93
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 QGIISSVVPPIHMETMVCVRYFNIRPLIVPGIKTGLNKKVD--NPREIGSDRIYNAV 115
QY 94 SAKFYGNVNVVISAGTALVDLVLE-GKFKGGFITLGLGKKILKLSDLAEGIPFFPEE 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 AASEYGPVIVDFGTATTCYIDSGVYGGALPGIMISTEALYNRAKLPRVDIAE 175
QY 153 VEIFLGRSTRCVLGAGYRESTEFIKSTLKLWRKFKVKKVITGGEGKYF-----SKFG 208
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 SSQIIGKTSVSSMQAGIFVFGQCEGIIAEMKKQSNAPVVVATGGLARMIKSSAVD 235
QY 209 IYDPLLVHRGMRNLLY 224
Db ||| : : : : : : : : : : : : : : : : : : : : : : : :
236 ILDPFLTKGL-ELLY 250

RESULT 5
ID Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilliales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-CLIP 11262 / SEROVAR 6a;
RA PubMed-11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusnlok C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst J., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596164; CAC95486.1; -
DR ListiList; LIN00253; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
```

```
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 12.8%; Score 149.5; DB 16; Length 259;
Best Local Similarity 23.4%; Pred. No. 0.00024;
Matches 60; Conservative 40; Mismatches 113; Indels 43; Gaps 7;

QY 4 LTVDVGNSSVDIALWEKKVKDFLKLKLSHEEFKKEFPKALKALGISVKQSFSEK----- 56
Db 1:||||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 LVIVIDGNTNCTGVYKEQKL-----LKHWRMTDRHRTSDGLGTMVLFNFFSYANLTPSDI 57
QY 57 ----VRGKIPKIKELK-----KENFPIDVDYKTPETLGTDRVALAY 93
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 QGIISSVVPPIHMETMVCVRYFNIRPLIVPGIKTGLNKKVD--NPREIGSDRIYNAV 115
QY 94 SAKFYGNVNVVISAGTALVDLVLE-GKFKGGFITLGLGKKILKLSDLAEGIPFFPEE 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 AASEYGPVIVDFGTATTCYIDSGVYGGALPGIMISTEALYNRAKLPRVDIAE 175
QY 153 VEIFLGRSTRCVLGAGYRESTEFIKSTLKLWRKFKVKKVITGGEGKYF-----SKFG 208
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 SSQIIGKTSVSSMQAGIFVFGQCEGIIAEMKKQSNAPVVVATGGLARMIKSSAVD 235
QY 209 IYDPLLVHRGMRNLLY 224
Db ||| : : : : : : : : : : : : : : : : : : : : : : : :
236 ILDPFLTKGL-ELLY 250

RESULT 6
ID Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RX MEDLINE-21886394; PubMed-11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusck G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010586; AAL94957.1; -
KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 12.4%; Score 145; DB 16; Length 256;
Best Local Similarity 22.0%; Pred. No. 0.00054;
Matches 55; Conservative 49; Mismatches 116; Indels 30; Gaps 6;

QY 6 VDVGNSSVDIALWE--GKKVKDFLKLKLSHEEFKKEFPK-----LKALGISVKQSFSEKVR 58
Db 1:||||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 IDIGNTHIVTGIYDNGNELISFRIATNDKMTDEYFSYFNNTKYNEISIKKVDAILIS 64
QY 59 GKIPKI-----KFLKENFPIDVDYKTPET-----LGTDRVA-LAYSAKKFY 99
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 SVYPNIITFQPFARKYKFEATIVDLKKLPFFAKNGINTGTGFGADRIDITEAMQKYP 124
QY 100 KNVVVVISAGTALVDLVLEGGKFGFITLGLGKKILKLSDLAEGIPFFPEEVEIFLGR 159
Db ||| : : : : : : : : : : : : : : : : : : : : : : : :
125 DNKLVIQFGTATTYDLKKGVYIGGILPGIDMSINALYNTAKLPRVKVFTTPSSVLGT 184
```









Search completed: June 24, 2003, 21:59:28  
Job time : 19.6773 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 3.98294 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453a-12

Perfect score: 1168

Sequence: 1 MRLFTVDVGNSSVDIALWEG.....YDPLLVHGRMNLILYHRI 229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

---Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	10.0	233	1 YACB_BACSU	P37564 bacillus su
2	101	8.6	596	1 UUP_BUCAI	P57445 buchera ap
3	100.5	8.6	733	1 K6AA_XENLA	P10665 xenopus lae
4	98	8.4	491	1 RNL_HAEIN	P45175 haemophilus
5	96	8.2	390	1 LA_DROME	P40796 drosophila
6	92	7.9	194	1 TRPG_HELPJ	Q32ju6 helicobacte
7	91	7.8	265	1 MOVV_CRMV	Q66221 chinese rap
8	91	7.8	463	1 NAOX_METJA	Q58065 methanococc
9	90.5	7.7	379	1 YOIL_MOUSE	P11260 mus musculus
10	90	7.7	370	1 DCUP_SCHPO	Q9usj5 schizosacch
11	89	7.6	799	1 HIS2_YEAST	P00815 saccharomyc
12	88.5	7.6	480	1 SYFA_METJA	Q57911 methanococc
13	88	7.5	213	1 RS6E_SULTO	Q975n7 sulfolobus
14	87	7.4	397	1 KAPI_BOVIN	P24256 bos taurus
15	87	7.4	593	1 CC23_SCHPO	Q42709 schizosacch
16	86.5	7.4	194	1 TRPG_HELPY	Q35868 helicobacte
17	86.5	7.4	629	1 K6AB_XENLA	P10666 xenopus lae
18	86.5	7.4	630	1 YO19_METJA	Q60325 methanococc
19	86.5	7.4	708	1 EFLG_STRCO	P40173 streptomyce
20	86.5	7.4	1176	1 SIAP_BACSH	P38537 bacillus sp
21	86	7.4	265	1 FANB_AQUAE	Q67783 aquifex aeo
22	86	7.4	297	1 MURB_AQUAE	Q68605 aquifex aeo
23	86	7.4	461	1 HMCS_ARATH	P54873 arabidopsis
24	85.5	7.3	326	1 SPI_BACBR	P43131 bacillus br
25	85.5	7.3	350	1 KAPA_MOUSE	P05132 mus musculus
26	85	7.3	483	1 MRCV_BUCAP	Q51926 buchera ap
27	84	7.2	267	1 MOVV_TVCV	Q88921 turnip vein
28	84	7.2	340	1 RECA_THEQA	P48296 thermus aqu
29	84	7.2	621	1 GIDA_RICPR	Q9ze90 rickettsia
30	84	7.2	976	1 PDR3_YEAST	P33200 saccharomyc
31	83.5	7.1	248	1 YAI4_SCHPO	Q09686 schizosacch
32	83.5	7.1	441	1 DCTA_RHIME	P20672 rhizobium m
33	83	7.1	144	1 HBAL_ONCMY	P02019 oncorhynch

34	82.5	7.1	848	1 AMPN_LACLA	Q48656 lactococcus
35	82	7.0	251	1 Y7SR_METFE	P21336 methanococ
36	82	7.0	448	1 Y861_METJA	Q58271 thermococc
37	82	7.0	546	1 THSB_THEKL	Q24730 thermococc
38	81.5	7.0	255	1 LACR_LACLA	P18816 lactococcus
39	81.5	7.0	350	1 KAPB_BOVIN	P05131 bos taurus
40	81.5	7.0	350	1 KAPB_HUMAN	P22894 homo sapien
41	81.5	7.0	358	1 IDI2_LISMO	Q8yfa5 listeria mo
42	81.5	7.0	1019	1 IDE_MOUSE	Q9jhr7 mus musculu
43	81.5	7.0	1019	1 IDE_RAT	P35559 rattus norv
44	81	6.9	365	1 Y075_METJA	Q60388 methanococ
45	81	6.9	546	1 THSB_PYRKO	Q52500 pyrococcus

#### ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_FACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowska A., Seror S.J., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,  
Viari A., Wambuy R., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





```

RX MEDLINE-94309661; PubMed-8035818;
RA Yoo C.J., Wolin S.L.;
RT "La proteins from Drosophila melanogaster and Saccharomyces
RT cerevisiae: a yeast homolog of the La autoantigen is dispensable for
RT growth."
RL Mol. Cell. Biol. 14:5412-5424(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwu K.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strung R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J.J., Seh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTION TERMINATION BY RNA
CC POLYMERASE III. BINDS RNA AND DNA. BINDS TO PRECURSORS OF RNA
CC POLYMERASE III TRANSCRIPTS. MAY PLAY A SPECIALIZED ROLE DURING FLY
CC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL,
CC PUPAL, AND ADULT DEVELOPMENT. EXPRESSION THROUGHOUT THE EMBRYO IS
CC FOLLOWED BY A RESTRICTED PATTERN OF MESODERMAL EXPRESSION THAT IS
CC LATER CONFINED TO THE VISCERAL MESODERM, GONADS, GUT, AND SALIVARY
CC GLANDS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: TO VERTEBRATE PROTEIN LA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U07652; AAA20518.1;
CC EMBL; L32988; AAA21776.1;
CC EMBL; AE003666; AAF53885.1;
CC FlyBase; FBgn0011638; La.
DR InterPro: IPR002344; Lupus La.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00302; LUPUSLA.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW RNA-binding; Nuclear protein; DNA-binding.
FT DOMAIN 149 234 RNA-BINDING (RRM).
FT CONFLICT 169 169 A -> T (IN REF. 1).
FT CONFLICT 183 183 KH -> NS (IN REF. 1).
FT CONFLICT 283 283 A -> R (IN REF. 1).
FT CONFLICT 329 329 K -> N (IN REF. 1).
SQ SEQUENCE 390 AA; 44884 MW; A8099288B90446A5 CRC64;

Query Match 8.2%; Score 96; DB 1; Length 390;
Best Local Similarity 22.8%; Pred. No. 1.8;
Matches 46; Conservative 31; Mismatches 45; Indels 80; Gaps 10;

QY 25 DFLKLSHEEFLKE-----EPPKLKALGI---SVKQSFSEKVRGKIPKIK 65
DB 230 DYLKKEQEEYAKNKKRKKKEAKPEPAFLPK-NAIVFEGAPETSSREIREAFKIK 288
QY 66 FLKKENFPI-QVDYKTPETLGTORVALYSAKFFYGNVWVISAGTALVDLVLEGKFKG 124
DB 289 -----DFEYVYIEFAKGETGKSVRLTADAEKVIK-----VEEGKLK- 327
QY 125 GFTTLGLGKKLKLSDLAEGIPFEFFPEEVEIFGRSTRECVLGGAYRESTEFISTLKLW 184
DB 328 -----FKDEVSLSLRKATEE-----EKEFEFIDKAEIFM 355
QY 185 RKVFKRKKFKVITGGEGKVFESK 206
DB 356 KK--RRDF-----TRNKGKFRNR 371

RESULT 6
TRPG_HELPJ STANDARD; PRT; 194 AA.
AC Q9ZJU6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Anthranilate synthase component II (EC 4.1.3.27) (Glutamine amido-
DE transferase).
GN TRPG OR JHPI202.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99120557; PubMed-9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: Chorismate + L-glutamine -> anthranilate +
CC pyruvate + L-glutamate.
CC -1- PATHWAY: Tryptophan biosynthesis; first step.
CC -1- SUBUNIT: Tetramer of two components I and two components II.
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE
CC USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES
CC GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001546; RAD06781.1; -.
CC HSP; Q06129; IQDL.
CC InterPro; IPR002385; Anth_synthII.
CC InterPro; IPR001317; CPS_GATase.
CC InterPro; IPR000991; GATase_1.
CC Pfam; PF00117; GATase; 1.
CC PRINTS; PR00097; ANTSYNTHASEII.
CC PRINTS; PR00099; CPSGATASE.
CC PRINTS; PR00096; GATASE.
CC TIGRFAMS; TIGR00566; trpG_papa; 1.
CC PROSITE; PS00442; GATASE_TYPE_I; 1.
CC TrypTophan biosynthesis; Lyase; Glutamine amidotransferase;
CC Complete proteome.
CC ACT_SITE 84 84 BY SIMILARITY.
CC ACT_SITE 170 170 BY SIMILARITY.
CC ACT_SITE 172 172 BY SIMILARITY.
CC ACT_SITE 194 194 BY SIMILARITY.
CC SEQUENCE 194 AA; 21479 MW; 47A123C678DBE119 CRC64;
CC -----
CC Query Match 7.9%; Score 92; DB 1; Length 194;
CC Best Local Similarity 26.8%; Pred. No. 1.7;
CC Matches 42; Conservative 18; Mismatches 63; Indels 34; Gaps 6;
CC -----
CC QY 35 LKEEFFPKALGIVSKQSFSEKVGKPKIKELKKENFPIQVDYKTPETIGTRVALAYS 94
CC DB 42 LMNEEKPTELPLFISPPG-PNPSSGNLLKIIMAKKKFPI-----LGVCLGLQALQASVG 95
CC QY 95 AKFYGKNNVVVISAGTALVIDLVLEGFKG-----GFTILGLKKLKILSLAE 143
CC DB 96 AKIIRSKFIV---HGKATTIATLAKHAFVKGGLGESVVGRYHSLMASGLPNLEVIAE-HD 151
CC QY 144 GIP-----EFPFEVEIFLGRSTRECVLG 167
CC DB 152 NIPMAIVNEEDKILAYOFHPESIMTLQGRALLEQSVG 188
CC -----
CC RESULT 7
CC MOV_P_CRMV STANDARD; PRT; 265 AA.
CC ID Q06221;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Movement protein (Cell-to-cell transport protein) (30 kDa protein).
CC GN MP.
CC OS Chinese rape mosaic virus (CRMV) (Oilseed rape mosaic virus).
CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
CC RN NCBI_TaxID=42007;
CC RX [1]
CC RP SEQUENCE FROM N.A.
CC RA MEDLINE=96197410; PubMed=8616237;
CC RA Aguilar I., Sanchez F., Martin-Martin A., Martinez-Herrera D.,
CC RA Ponz F.;
CC RT "Nucleotide sequence of Chinese rape mosaic virus (oilseed rape mosaic
CC virus), a crucifer tobamovirus infectious on Arabidopsis thaliana.";
CC RL Plant Mol. Biol. 30:191-197(1996).
CC CC -1- FUNCTION: Involved in transport of the virus from the initially
CC infected cells to adjacent cells, possibly by modifying the
CC function of the plasmodesmata. Also influences local lesion
CC development. Binds to RNA and single-stranded DNA.
CC CC -1- SIMILARITY: BELONGS TO THE TOBAMOVIRUSES MOVEMENT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

DR EMBL: U67512; AAB98641.1; --  
 DR TIGR: MJ0649; --  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR004099; pyr\_redox\_dim.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR Pfam: PF02852; pyr\_redox\_dim; 1.  
 DR ProDom: PD0001139; FAD\_pyr\_redox; 1.  
 KW Hypothetical protein; Oxidoreductase; NAD; Flavoprotein; FAD;  
 KW Redox-active center; Complete proteome.  
 FT ACT\_SITE 57 REDOX-ACTIVE.  
 FT NP\_BIND 167 182 NAD (BY SIMILARITY).  
 FT NP\_BIND 286 296 FAD (FLAVIN PART) (BY SIMILARITY).  
 SQ SEQUENCE 463 AA; 50599 MW; AA216A9F20E6C1AA CRC64;

Query Match  
 Best Local Similarity 7.8%; Score 91; DB 1; Length 463;  
 Matches 48; Conservative 41; Mismatches 79; Indels 66; Gaps 10;

QY 19 EGKVKDFLKLSHEE-----FLKEEFKPKKAL-----GISVKQSPSEKVRG----- 59  
 DB 136 EGKDLGVFKFTIEDGRAILKYEENGCKKVAVVGAGAGIGLEWAGYGLKRCGLDLVLWEM 195  
 QY 60 ---KIPKI-----FKLKENFPIQVDYKTPETLGTDRVALAYSAKKFKYKKNVVI 106  
 DB 196 AQVLFRLDPDMAEIVQILEKEGKIVMLSKPLEKIVGKVEYVYVDGKLYDVMVIM 255  
 QY 107 SAGTALVIDLVLEGKFGKGGFITLGLGKILKILSDLAEGIPPEFPEVEIFLGRSTRECVL 166  
 DB 256 ATGVRNIELAKGCKGKIGFAIEVNEKMQ-----TSPNIV-----AVGDCV- 298  
 QY 167 GGAYRESTFEI--KSTLKLWRKFKRKVKV-----ITGGEKGYF-----SKFG 208  
 DB 299 -----EVIDFITGKTLSPFGTAARQGVAGKNIAGVEAKFPVLNSAVSKIG 347

RESULT 9  
 Y01L\_MOUSE STANDARD; PRT; 379 AA.  
 AC P11260;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-JUL-1989 (Rel. 11, Last annotation update)  
 DE Hypothetical protein ORF-1137.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 SEQUENCE FROM N.A.  
 X MEDLINE=87064284; PubMed=3023821;  
 RA Loeb D.D., Padgett R.W., Hardies S.C., Shehee W.R., Comer M.B.,  
 RA Egell M.H., Hutchison C.A. III;  
 RT "The sequence of a large L1MD element reveals a tandemly repeated 5'  
 RT end and several features found in retrotransposons.";  
 RL Mol. Cell. Biol. 6:168-182(1986).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M13002; AAG66023.1; ALT\_INIT.  
 DR PIR: A24906; QOMSL.  
 DR InterPro: IPR004244; Transposase\_22.  
 DR Pfam: PF02994; Transposase\_22; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 379 AA; 43788 MW; F4DDE312807644D3 CRC64;

Query Match  
 7.7%; Score 90.5; DB 1; Length 379;

Best Local Similarity 23.5%; Pred. No. 4.7;  
 Matches 51; Conservative 38; Mismatches 73; Indels 55; Gaps 12;

QY 12 SVDIALWEGKVKDFLKLSHEEFLKEEF--PKLKALGISVKQSPSEKVRGKIPKIFLK 69  
 DB 171 NIDTYKENTKCKRIL-TQNIQVQIDTPMRNRLIIGIDENEDF--QLKGPANTFNK 227  
 QY 70 ENFP-----IQVDYKTPETLGTDRVALAYSAKKFKYKKNVVISAGTALVIDLV 120  
 DB 228 ENFPNKKEMPMIIQEAERTPNRLDQKRNRRH-----IIRTNALNKDRILKA 277  
 QY 121 KFKGGFTLGLGKILKILSDLAEGIPPEFPEVEIFLGRSTRECVLGAGYRESTEF 180  
 DB 278 VREKGVQTY-KGRPIRI-----TPDFSPETMK-----ARRAWTDVIO-T 314  
 QY 181 LKLW-----RKVEKRRKFKVWITGGGKVF--SKFGIY 210  
 DB 315 LREHKQCPRLLYPAKLSITI-DGETKVFHDKTKFTQY 350

RESULT 10  
 DCUP\_SCHPO STANDARD; PRT; 370 AA.  
 AC Q9USJ5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Uroporphyrinogen decarboxylase (EC 4.1.1.37) (URO-D) (UPD).  
 GN SPC4483.05C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feilwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Muncall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -I- CATALYTIC ACTIVITY: Uroporphyrinogen-III = coproporphyrinogen + 4  
 CC CO(2).  
 CC -I- PATHWAY: PORPHYRIN AND HEME BIOSYNTHESIS.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -I- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AL132870; CAB60679.1; -  
 DR HSP; P06132; LURO.  
 DR InterPro; IPR000257; Uro\_decabxyls.  
 DR Pfam; PF01208; URO-D; 1.  
 DR PROSITE; PS00906; UROD.1; FALSE\_NEG.  
 DR PROSITE; PS00907; UROD.2; 1.  
 KW: Lyase; Decarboxylase; Porphyrin biosynthesis; Heme biosynthesis.  
 SQ SEQUENCE 370 AA; 41846 MW; 110DFC463733FD89 CRC64;

Query Match 7.7%; Score 90; DB 1; Length 370;

Best Local Similarity 28.0%; Pred. No. 5;

Matches 42; Conservative 20; Mismatches 66; Indels 22; Gaps 6;

18 WEGKVKDFKLISHEEFKEPPKALGIVSKQSFSEKVRGKIPKIFLKKENFPI-QV 76

224 WAG-----ELSPEDFTEYAPYLVLRICQEVKQHLKKKKRDEVPMIVFAKAWAYIDQL 276

77 DYKPTCLTDRVALAYSAKFYGNVV-----ISAGTALVIDLVLEKFKGGFTT 128

277 CDSGYDVLGLDWTSPKAVRIGRNVTFQGNLDPNLYLTREII-----EARTREMQD 332

129 LGLGKKLILSDLAGI-PEFFPPEVEIFL 157

333 FGGGKGQYII-NLHGITPGVNPDDVREFL 361

# RESULT 11

## HIS2\_YEAST

ID HIS2\_YEAST STANDARD; PRT; 799 AA.

AC P00815;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAY-1992 (Rel. 22; Last sequence update)

DT 15-JUN-2002 (Rel. 41; Last annotation update)

DE Histidine biosynthesis trifunctional protein [Includes:

DE Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19); Phosphoribosyl-ATP

DE pyrophosphohydrolase (EC 3.6.1.31); Histidinol dehydrogenase

DE (EC 1.1.1.23) (HDB)]

GN HIS4 OR YCL030C OR YCL30C OR YCL183.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91377317; PubMed=1897318;

RA Rad M.R., Luetzenkirchen K., Xu G., Kleinhans U., Hollenberg C.P.;

RT "The complete sequence of a 11,953 bp fragment from CIG on chromosome

RII encompasses four new open reading frames.";

RL Yeast 7:533-538(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=82262805; PubMed=7049842;

RA Donahue T.F., Farabaugh P.J., Fink G.R.;

RT "The nucleotide sequence of the HIS4 region of yeast."

RL Gene 18:47-59(1982).

RN [3]

RP SEQUENCE OF 1-20 FROM N.A.

RX MEDLINE=80254613; PubMed=6250062;

RA Farabaugh P.J., Fink G.R.;

RT "Insertion of the eukaryotic transposable element Ty1 creates a

RI 5-base pair duplication."

RL Nature 286:352-356(1980).

RN [4]

RP SEQUENCE OF 1-16 FROM N.A.

RX MEDLINE=85270536; PubMed=2991923;

RA Roeder G.S., Rose A.B., Pearlman R.E.;

RT "Transposable element sequences involved in the enhancement of yeast

RI gene expression."

RL

Proc. Natl. Acad. Sci. U.S.A. 82:5428-5432(1985).  
 -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-ATP + H(2)O -> 5-phosphoribosyl-AMP + diphosphate.  
 -!- CATALYTIC ACTIVITY: 5-phosphoribosyl-AMP + H(2)O -> 5-(5-phosphoribosylaminoformimino)-1-(5-phosphoribosyl) imidazole-4-carboxamide.  
 -!- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O -> L-histidine + 2 NADH.  
 -!- PATHWAY: Histidine biosynthesis; second step.  
 -!- PATHWAY: Histidine biosynthesis; third step.  
 -!- PATHWAY: Histidine biosynthesis; ninth (last) step.  
 -!- SIMILARITY: IN THE HDH DOMAIN, TO OTHER PROKARYOTIC, FUNGAL AND PLANTS HDH.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; V01309; CAA24616.1; -  
 DR EMBL; V01310; CAA24617.1; -  
 DR EMBL; J01331; AAA88722.1; -  
 DR EMBL; X59720; CAA42335.1; -  
 DR EMBL; M11491; AAA67504.1; -  
 DR EMBL; M11492; AAA67505.1; -  
 DR EMBL; M11694; AAA18400.1; -  
 DR EMBL; M11695; AAA18401.1; -  
 DR EMBL; M11696; AAA18402.1; -  
 DR PIR; S17473; SHBY.  
 DR SGD; S0000535; HIS4.

DR InterPro; IPR001692; Histidinol\_dh.

DR InterPro; IPR002496; PRA-CH.

DR InterPro; IPR002497; PRA-PH.

DR Pfam; PF00815; Histidinol\_dh; 1.

DR Pfam; PF01502; PRA-CH; 1.

DR PRINTS; PR00083; HOLHDSRGNASE.

DR ProDom; PD002610; PRA-CH; 1.

DR ProDom; PD002611; PRA-PH; 1.

DR ProDom; PD002680; Histidinol\_dh; 1.

DR TIGRFAMs; TIGR00069; hisD; 1.

DR PROSITE; PS00611; HISOL\_DEHYDROGENASE; 1.

KW Histidine biosynthesis; Multifunctional enzyme; Hydrolase;

KW Oxidoreductase; NAD.

FT DOMAIN 1 229 PHOSPHORIBOSYL-AMP CYCLOHYDROLASE.

FT DOMAIN 230 312 PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE.

FT ACT\_SITE 470 470 HISTIDINOL DEHYDROGENASE.

FT CONFLICT 53 53 BY SIMILARITY.

FT CONFLICT 375 375 A -> R (IN REF. 2).

FT CONFLICT 375 375 R -> S (IN REF. 2).

FT CONFLICT 385 385 H -> Y (IN REF. 2).

FT CONFLICT 402 403 AL -> VF (IN REF. 2).

FT CONFLICT 441 441 D -> N (IN REF. 2).

FT CONFLICT 794 794 I -> F (IN REF. 2).

SQ SEQUENCE 799 AA; 87790 MW; AA154CD7E032409E CRC64;

Query Match 7.6%; Score 89; DB 1; Length 799;

Best Local Similarity 23.9%; Pred. No. 14;

Matches 63; Conservative 23; Mismatches 62; Indels 116; Gaps 14;

QY 14 DIALWEGKK-----VKDFLKSHEE---FLKEEFPLKAL----- 45

DB 11 DIASWNSKKEYSLVGVLLDGSLSNEEILQFSKEEVLVALSLPSKFSDEIIAFL 70

QY 46 --GIS-----VKDFLKSHEE---FLKEEFPLKAL----- 45

DB 71 NNGVSSLFTASQDAKTAETHLVEQLNVPKRVVEENGVSQNFVKKQFSQ---DKIVSI 127

QY 65 KFLKKENFFIQVDYKTPETIG---TDR-----VALAYSAKFKYGNVVV 105

```
Db 128 KLSK-----DMLTKEVLGEVTRDRPGLYITLVVDQYERCLGLVYSKSKSIK---A 177
QY 106 ISAGTALVI-----DIVLEGKFGGTTTGLGKK-LKILSDLAEGIPPEPVEVEI-FLG 158
Db 178 IDLGRGVYYSRSRNEIWKGE-----TSNGQKLLQISTDCSDALKFIVEQENGVFCH 231
QY 159 RSTRECVLGGAYRESTEFKSTLK 182
Db 232 LETMSCF--GEFKHGLVGLSESLK 253

RESULT 12
SYFA_METJA
ID SYFA_METJA STANDARD; PRT; 480 AA.
AC Q57911;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20) (Phenylalanine-
  tRNA ligase alpha chain) (Phers).
IN PHES OR MJ0487.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fudown P.W., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
  jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
  diphosphate + L-phenylalanyl-tRNA(Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
  similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC PHE-TRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67499; AAB98478.1; -
CC HSSP: P27001; 1PTS.
CC TIGR: MJ0487; -
CC InterPro: IPR002106; AATRNA_ligaseII.
CC InterPro: IPR004529; PheS.
CC InterPro: IPR002319; tRNA-synt_2d.
CC Pfam: PF01409; tRNA-synt_2d; 1.
CC TIGRFAMs: TIGR00468; pheS; 1.
CC PROSITE: PS50862; AA_TRNA_LIGASE_II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
  Complete proteome.
SQ SEQUENCE 480 AA; 57027 MW; 1DC2599CF0D97B72 CRC64;

Query Match 7.6%; Score 88.5; DB 1; Length 480;
Best Local Similarity 20.9%; Pred. No. 8.8;
Matches 49; Conservative 42; Mismatches 75; Indels 69; Gaps 11;

Db 128 KLSK-----DMLTKEVLGEVTRDRPGLYITLVVDQYERCLGLVYSKSKSIK---A 177
QY 106 ISAGTALVI-----DIVLEGKFGGTTTGLGKK-LKILSDLAEGIPPEPVEVEI-FLG 158
Db 178 IDLGRGVYYSRSRNEIWKGE-----TSNGQKLLQISTDCSDALKFIVEQENGVFCH 231
QY 159 RSTRECVLGGAYRESTEFKSTLK 182
Db 232 LETMSCF--GEFKHGLVGLSESLK 253

RESULT 13
RS6E_SULTO
ID RS6E_SULTO STANDARD; PRT; 213 AA.
AC Q975N7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 13-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6e.
DE RPS6E OR ST0382.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Akai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RA "Complete genome sequence of an aerobic thermoacidophilic
  Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res. 8:123-140(2001).
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF000982; BAB65363.1; ALT_INIT.
CC InterPro: IPR001377; Ribosomal_S6e.
CC Pfam: PF01092; Ribosomal_S6e; 1.
CC ProDom: PD003460; Ribosomal_S6e; 1.
CC PROSITE: PS00578; RIBOSOMAL_S6e; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 213 AA; 23636 MW; 90FAL5FB3E3F36FF CRC64;

Query Match 7.5%; Score 88; DB 1; Length 213;
Best Local Similarity 26.3%; Pred. No. 3.8;
Matches 46; Conservative 20; Mismatches 43; Indels 66; Gaps 10;

QY 19 EGKKVKDFLKLSEELKEEFPKLKALGI-----SVKQSFSEKVGKPKIKFLKKN 71
Db 37 EGKAL-PIAKINOK--LKEE-----LGLTLLTLQTIKQEGDKVKV-----VKTH 78
QY 72 FPIQVDYKTPETLGTDRVALAYSARKFYGNVV-----VISAGT 110
Db 79 FKIEIDNNVP-----SDEWTAASIAEKYGANEFAYRTKSPOLSDQSKLSNIIGAKI 134
```

[illegible]

SEQUENCE FROM N.A.  
RP STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkheart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.35258 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-12  
Perfect score: 1168  
Sequence: 1 MRELTVDVGNSSVDIALWEG.....YDPLLVRGMRNLLYLYHRI 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*

- 1: pir1:\*
- 2: pir2:\*
- 3: pir3:\*
- 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES.

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	100.0	229	E70465	hypothetical prote
2	193.5	16.6	246	D72320	conserved hypotet
3	177.5	15.2	209	H81382	hypothetical prote
4	156.5	13.4	259	AF1102	conserved hypotet
5	149.5	12.8	259	AF1464	conserved hypotet
6	141.5	12.1	262	F70165	conserved hypotet
7	138.5	11.9	223	F64627	hypothetical prote
8	138.5	11.9	254	F83660	hypothetical prote
9	138.5	11.9	273	E97293	probable transcrip
10	133.5	11.4	262	E75516	conserved hypotet
11	133	11.4	257	S75559	hypothetical prote
12	131.5	11.3	223	G74887	hypothetical prote
13	129.5	11.1	261	B87489	transcription acti
14	128.5	11.0	265	T36391	hypothetical prote
15	122.5	10.5	273	D71326	conserved hypotet
16	117	10.0	233	S66100	conserved hypotet
17	117	10.0	276	A12292	hypothetical prote
18	114	9.8	272	A70955	hypothetical prote
19	112	9.6	274	H86937	conserved hypotet
20	108	9.2	592	D82031	probable biotin-fa
21	105	9.0	224	A99571	conserved hypotet
22	103	8.8	592	B81009	Birk protein/Bvg a
23	102.5	8.8	254	G69495	transcription regu
24	101	8.6	596	D84972	ABC transporter At
25	100.5	8.6	733	B30001	ribosomal protein
26	98.5	8.4	248	H83111	hypothetical prote
27	98	8.4	390	A53773	La/SS-B homolog D
28	98	8.4	491	F64118	cytosolic axial fi
29	97.5	8.3	484	F71954	guanosine-5'-triph

30 96 8.2 390 2 A53781 ribonucleoprotein  
31 96 8.2 578 2 C97167 uncharacterized co  
32 95.5 8.2 426 1 B71249 hypothetical prote  
33 95 8.1 242 2 A82637 conserved hypotet  
34 94.5 8.1 484 2 F64554 guanosine pentapho  
35 93.5 8.0 427 2 T39113 probable flavohemo  
36 92 7.9 194 2 E71836 antranilate synth  
37 91 7.8 285 2 S65054 movement protein -  
38 91 7.8 463 2 A64381 NADH oxidase - Met  
39 90.5 7.7 235 2 H81451 ubiquinone/menaqui  
40 90.5 7.7 357 1 QOMSIL retrovirus-related  
41 90 7.7 370 2 T50443 probable uroporphy  
42 89 7.6 286 2 B70430 transcription regu  
43 89 7.6 799 1 SHBY phosphoribosyl-AMP  
44 88.5 7.6 462 2 D85438 hypothetical prote  
45 88.5 7.6 470 2 T45962 hypothetical prote

ALIGNMENTS

RESULT 1  
E70465  
hypothetical protein aq\_1924 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: E70465  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70465  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-229 <Q>  
A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AC07720.1; PID:g2984188; GB:AE  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aq\_1924

Query Match 100.0%; Score 1168; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. NO. 2.6e-82;  
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRELTVDVGNSSVDIALWEGKVKDFLKSHEEFKKEFPKLGKALGTSVKQSFSEKVRGK 60  
|||||  
Db 1 MRELTVDVGNSSVDIALWEGKVKDFLKSHEEFKKEFPKLGKALGTSVKQSFSEKVRGK 60  
|||||  
QY 61 IPKIFLKKENFFIQVDYKTPETLGTDRVALAYSAKKFKYGNVYVISAGTALVIDLVLEG 120  
|||||  
Db 61 IPKIFLKKENFFIQVDYKTPETLGTDRVALAYSAKKFKYGNVYVISAGTALVIDLVLEG 120  
|||||  
QY 121 KFGGFTTGLGKKILSLDLAEGIPFPEVEIFLGRSTRECVLGAYRESEFFIKST 180  
|||||  
Db 121 KFGGFTTGLGKKILSLDLAEGIPFPEVEIFLGRSTRECVLGAYRESEFFIKST 180  
|||||  
QY 181 LKLWRKVRKFRKFWITGGEGYFSKFGIYDPLLVRGMRNLLYLYHRI 229  
|||||  
Db 181 LKLWRKVRKFRKFWITGGEGYFSKFGIYDPLLVRGMRNLLYLYHRI 229  
|||||

RESULT 2  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain.MS8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: D72320  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: D72320

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-246 <ARN>

A:Cross-references: GB:AE001754; GB:AE000512; MID:g4981417; PIDN:AAD35964.1; PID:g4981417

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0883

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 16.6%; Score 193.5; DB 2; Length 246;  
Best Local Similarity 26.5%; Pred. No. 1.2e-07;  
Matches 66; Conservative 45; Mismatches 105; Indels 33; Gaps 6;

QY 3 FLVDVGNSSVDIALWEGKVKDKFLKLSHEEFLKEE-----PPKALKGI-- 47

DB 2 YLLVDVGNTHSVFSTEDGKTRRRRLSTGVFQTEDELFSLHPLLDGAMREIKGIGVAS 61

QY 48 -----SVQSFSEKVRGKPKPKFKKENPFDVYKTPETLGTDRVALAYSAKFPYK 101

JB 62 VVPTQNTVIERFSQKVFHISP--IWKAKNGCCVKNVKNPSEVGADRVANVAVFVKEYGK 119

QY 102 NVVVISAGTALVDLVLEGKFGGFTVLGLGKKILSDLAEGIPPEFPEVEIFLGRST 161

DB 120 NGIIIDMGATTVDLVNVSYEGAILPGFFMMVHSLFRGTAKPLPLVEKPADFVVGKDT 179

QY 162 RECVLGAGVRESEFTKSTLKLWRKFKKVVITGGEGKYFS---KFGIYDPLLVHRG 218

DB 180 EENIRLVGNVSVYALEGIIGRIKEVY-GDLPVLTGGQSKIIVKDMIKHEIFEDLTIGK 238

QY 219 MRNLLLYH 227

DB 239 -----VYH 241

RESULT 3

H81382

hypothetical protein Cj0394c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: H81382

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: H81382

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-209 <PAR>

A:Cross-references: GB:AL139075; GB:AL111168; MID:g6967817; PIDN:CAB74230.1; PID:g696786

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0394c

C:Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 15.2%; Score 177.5; DB 2; Length 209;

Best Local Similarity 25.8%; Pred. No. 1.6e-06;

Matches 62; Conservative 44; Mismatches 83; Indels 51; Gaps 10;

QY 4 FLVDVGNSSVDIALWEGK----KVKDFLKLSSHEEFLKEEPPKALKGISVQSFSEKVRG 59

DB 2 LLCIDIGNSANF-LDDNKVFTLNIDQFLFKNEQ-----KIFVNVNEHLKHLK- 50

QY 60 KPKIKFLKKNENPIQVDYKTPET----LCTDRVALAYSAKKFKYKNNVVVISAGTALVID 115

DB 51 -----NOKNFINLEPYFLFTIYVGLGIDRIINACYTI-----EDGVVVDAGSAITID 97

QY 116 LVLEKFKFGKFTVLGLGKKILSDLAEGIPPEFPEVEI-FLGRSTRECVLGGVREST 174

DB 98 IISNSHLGGFILPGIANYKKIYSHISPRLKSEFNTOVSLDAPFOKTMDSALSGVPKGIY 157

QY 175 EFIKSTLKLWRKFKRKVVITGGEGKYSKPF---GIYDPLLVHRGMR-----NLLY 224

DB 158 LLIKDA-----AQNKRLYFTGGDQQLFANYFDHAIYDKLLIFRGMKKIKENPNLLY 209

RESULT 4

AF1102

conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain F

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AF1102

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.

ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <GLA>

A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0221

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 13.4%; Score 156.5; DB 2; Length 259;

Best Local Similarity 23.8%; Pred. No. 8.3e-05;

Matches 61; Conservative 39; Mismatches 113; Indels 43; Gaps 7;

QY 4 LTVDVGNSSVDIALWEGKVKDKFLKLSHEEFLKEEPPKALKGISVQSFSEK----- 56

DB 3 LVLDVGNNTCTGVYKQKL-----LKHWRMTTDRHRTSDGLMTLVNFFSVANLTPSDI 57

QY 57 -----VRGKIPKIFLK-----KENFPIQVDYKTPETLGTDRVALAY 93

DB 58 QGIIISVVPPIHMETCMVCYRINRPIIVGPGIKTGLNLKVD--NPREIGSDRIVNAV 115

QY 94 SAKFKYKNNVVISAGTALVIDLVLE-GKFKGGFTVLGLGKKILSDLAEGIPPEFPEE 152

DB 116 AASEEYGPVIVDFGATTCYIDEGVYOGGAIAPGIMISTEALYNRAAKLPVDIAE 175

QY 153 VEIFLGRSTRECVLGGVAYRESEFTKSTLKLWRKFKKVVITGGEGKYF----SRFG 208

DB 176 SSQIIGKSTVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGLARMTTEKSSAVD 235

QY 209 IYDPLLVHRGMRNLLY 224

DB 236 ILDPFLLTKGL-ELLY 250

RESULT 5

AF1464

conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AF1464

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi

Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.

ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1464

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178





Db 123 YPAIVVDFGATTCYCLINEKKQYAGGVIAPIGIMISTEALYHRASKLPRIEIAKPKQVVG 182  
QY 160 STRECVLGGAYRSTEEIKSTLKLWRKVRKPKVVTGGGKVF-----SKFGYIDPLV 215  
Db 183 NTIDSMOSGIFYGVSOVDGVWRKMAQSEPKVIATGGLAKLIGTSETIDVIDSFLT 242  
QY 216 HRGMRNLLY 224  
Db 243 LKGLQ-LIY 250

RESULT 9  
E97293  
Probable transcription regulator, homolog of Bvg accessory factor. [imported] - Clostridium  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: E97293  
R: Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97293  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3200  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.9%; Score 138.5; DB 2; Length 273;  
Best Local Similarity 21.6%; Pred. No. 0.0021;  
Matches 58; Conservative 54; Mismatches 103; Indels 53; Gaps 11;

QY 4 LTVDVGNSSVDIALWEGKKVKDFLKLSSHEEFLEKEEPKPKALGIVSKQSFSEK-----56  
Db 14 LVLVDGNTINVLGIYNDKLTAEWRLSTDLRSAD-----EYGIQVNNLFQDDKLDPTLV 68

QY 57 -----VRGKIPKIF-----LKKNFP-----IQVDYKTPETLTGTRVALAYS 95  
Db 69 EGVLISSVVPNIMYSLEHMRKFKINPLVVGIGIKTINIKYDNKPEVGDRIVNAVA 128

QY 96 KFPYGNVVISAGTALVIDLLE-KGFKGGFTTLGLGKKLILSDLAEGIEPFFPEVE 154  
Db 129 HEIYRSLIIDFTATTCAVRENGDYLGGACPGIKVSSEALFEKAALP-----RVE 183

QY 155 IF-----LGRSTRECVLGG---AYRSTEFIKSTLK-LWRKVKRKKVVTGGGKYS 205  
Db 184 LKPAVACNKTISSTOSGIVGYIGQVRYIVRMKEELQEEGEKEPLVAVATGGLAKLIS 243

QY 206 K-----FGIYDPLLVHRGMRNLLYHRI 229  
Db 244 EEAKNVYDINPFLTEGLR-IIYEKNRV 270

RESULT 10  
E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WHI>

A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 11.4%; Score 133.5; DB 2; Length 262;  
Best Local Similarity 23.0%; Pred. No. 0.0049;  
Matches 58; Conservative 44; Mismatches 113; Indels 37; Gaps 9;

QY 4 LTVDVGNSSVDIALWE--GKKVKDFLKLSSHEEFLEKE-----FKP---LK 43  
Db 7 LAVDIGNTTVVLGLADASGALHTWRINTNREMLPDDLALQHLGTLTAGAPIPRAAVLS 66

QY 44 ALGISVKQSFSEKVRGKIPKIF-LKKENFP-IQVDYKTPETLTGTRVALAYS AKKFFY- 100  
Db 67 SVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRCLNLFGBAEKYLGG 126

QY 101 -KNVVVISAGTALVIDLLEK-KPKGGFTTLGLGKKLILSDLAEGIEPFFPEVEIFLG 158  
Db 127 LDYAVVYDFGTSTNFDVVGRRFLGGILATGAQVSADALFARAALPRITLQAPETAIG 186

QY 159 RSTRECVLGGAYRSTEFIKSTLKLRKVKRKKVVTGGGKYSF-----FGY 210  
Db 187 KNTVHALQSLVFGYAEWVDGLLRIRAEALPGEAVAVATGG-----FSRTVGOICQEIFDY 242

QY 211 DPLLVHRGMRNL 222  
Db 243 DETLTGRGLVEL 254

RESULT 11  
S75559  
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C:Accession: S75559  
R: Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O.; K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, M.; Ya  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S75559  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-257 <KAN>  
A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAAL18120.1; PID:d1c  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Start codon: GTG

Query Match 11.4%; Score 133; DB 2; Length 257;  
Best Local Similarity 28.3%; Pred. No. 0.0052;  
Matches 66; Conservative 31; Mismatches 118; Indels 18; Gaps 8;

QY 3 FLTVDVGNSSVDIALWEGKK-----VKDFLKLSSHEEFLEKEEPKPKALGIVSKQSFSEKV 57  
Db 20 WLGLMGNLRLHWAYCSGNAPLQWVTDYNPKSAQ--LPVLGLKVPMLASVVPQTEW 77

QY 58 RGKIPKIFLKKNFPIQVDYKTPETLTGTRVALAYS AKKFFYGNVVVISAGTALVI-DL 116  
Db 78 RVYQPKLTLK--NLPLVNLV---PSFGIDRALAGLTGLTYGFPCLVVDGDTALTITGF 132

QY 117 VLSEKFKGGFTTLGLGKKLILSDLAEGIEPFFPEVEIFLGR---STRECVLGGAYRES 173  
Db 133 DODKLVGGAILPGLGLQATLGLRLAALPKLEMDQLTELDRWALDTPPSAIFSGVYGV 192

QY 174 TEFIKSTLKLRKVKRKKVVTGGGKYSF-----GIYDPLLVHRGMRNLLY 225  
Db 193 LGALQSLQDWQKLFPCA-AMVITGGDGKILHGLFKEHSPNLSVANDONLIFL 244

Qy	53	--PSEYRGKIPKIIFLKKNFPQT-----VDYKTPTETLCTDRVALAYSAAKK	97
Dd	61	VIISSVVPQSIFNURNISRRFYNNPEVLPIGNAKLIGDIVREKPSEAAGADRLVNIGAAM	120
Qy	98	FYGNVVIVSAGTALVIDLV-LEGKFSGGITLGKGKKILSLDLAEGIPEFFPEEV--E	154
Dd	121	VYPCLPVIDGSTATTDFIVAADAFAFGGIAPGINLSQAALHEAAAALKPRATIORPAGN	180
Qy	155	IPLGRSTRECVELGGAYRESTEFISTLUKMRKYFKRKVKVITGTGEGYF-----SKFGIY	210
Dd	181	RIVGTDTVSAMOSGVFWGISLIIEGLVARIKAEGERPMVTIATGVGSALFEGETSIDHF	240
Qy	211	DPLLVRHGRMNLPLYLYHR	228
Dd	241	DSDTIRG---LDLEIYYR	255
RESULT 14			
T36391			
C:	hypothetical protein SCE94_3lc - Streptomyces coelicolor		
C:	Species: Streptomycetes coelicolor		
C:	Date:	03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000	
C:	Accession:	X6391	
R:	Olliver, K.; Harris,	D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandrea	
A:	submitted to the EMBL Data Library, April 1999		
A:	Reference number:	Z21573	
A:	Accession:	T36391	
A:	Status:	Preliminary; translated from GB/EMBL/DDBJ	
A:	Molecule type:	DNA	
A:	Residues:	1-265 <OLI>	
A:	Cross-references:	EMBL:AL049628; PIDB:CAB40880.1; GSPDB:GN00070; SCODEB:SCE9	
A:	Experimental source:	strain A3(2)	
C:	Genetics:		
A:	Gene:	SCE94_3lc	
C:	Superfamily:	Streptomycetes coelicolor hypothetical protein SCE94_3lc	
Query Match                  11.0%; Score 128.5; DB 2; Length 265;			
Best Local Similarity      24.1%; Pred. No. 0.012;			
Matches 49; Conservative 36; Mismatches 71; Indels 47; Gaps 9			
Qy	4	LTVDVGNSSVDIALWEKKKVDFLLKSHEELKEFPKLKA-----	45
Dd	3	LTIDVGNTHTVLGFDEGDIVEHWRISTD-SRTADELAVLQLGLMGHPHLGDELSDG	60
Qy	46	-GI-----SVKQSFSEKRV---GKTPKIKFLK--KENPFIOVDYKTPETLCGTRDAL	91
Dd	61	IDGAICATACPVSULHELREVTRYRYGDPVAVLPVEPGVKTGPILTTH--PKREGADRIN	118
Qy	92	AYSAAKPYGKNVVVISAGTALVIDLV-LEGKFSGGITLGKGKKILSLDLAEGIPEFFPE	150
Dd	119	AVAARELVGGPAIVDVFCTATTDFDAVSARGEYIGGVIAPIGBIEISREAL----GVKGAOL	173
Qy	151	EVEVI-----FLGRSTRCEVLGG	168
Dd	174	RKIEVARPSVIGKNTVEAMQG	196
RESULT 15			
D71326			
C:	conserved hypothetical protein TP0431 - syphilis spirochete		
C:	Species:	Treponema pallidum subsp. pallidum (syphilis spirochete)	
C:	Date:	24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000	
C:	Accession:	D71326	
R:	Fraser, C.M.; Norris,	S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson	
rson, J.; Khalak, H.; Richardson, D.; Howell, J.R.K.; Chidambaram, M.; Utterback			
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.			
Science 281, 375-388, 1998			
A:	Title:	Complete genome sequence of Treponema pallidum, the syphilis spiroche	
A:	Reference number:	A71250; MUID:98332770; PMID:9665876	
A:	Accession:	D71326	
A:	Status:	Preliminary; nucleic acid sequence not shown; translation not shown	
A:	Molecule type:	DNA	

Search completed: June 24, 2003, 22:03:11  
Job time : 10.3526 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 23.511 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-12

Perfect score: 1168

Sequence: 1 MRLFTVDVGVNSVDIALWEG.....YDPLLVHRGMNLLLYHRI 229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	100.0	229	AAU91159	Aquifex aeolicus p
2	225	19.3	257	AAU91174	Pantothenate kinas
3	193.5	16.6	246	AAU91156	Thermotoga maritim
4	177.5	15.2	209	AAU91165	Pantothenate kinas
5	167.5	14.3	262	AAU91170	Pantothenate kinas
6	164.5	14.1	241	AAU91179	Pantothenate kinas
7	158.5	13.6	258	AAU91153	Rhodobacter capsul
8	156.5	13.4	259	ABB47661	Listeria monocytog
9	154.5	13.2	258	AAU01243	B. subtilis novel
10	154.5	13.2	258	AAU91149	Bacillus subtilis

11	148.5	12.7	258	23	AAU91172	Pantothenate kinas
12	141.5	12.1	262	23	AAU91158	Borrelia burgdorfe
13	139.5	11.9	255	23	AAU91154	Geobacter sulfurre
14	138.5	11.9	223	19	AAW98422	H. pylori GHPO 344
15	138.5	11.9	223	23	AAU91161	Helicobacter pylor
16	138.5	11.9	223	23	AAU91181	Pantothenate kinas
17	138.5	11.9	254	23	AAU91171	Pantothenate kinas
18	133.5	11.4	262	23	AAU91155	Deinococcus radiop
19	133	11.4	256	23	AAU91175	Pantothenate kinas
20	133	11.4	257	23	AAU91168	Synechocystis pant
21	132.5	11.3	244	23	AAU91166	Pantothenate kinas
22	130	11.1	219	23	AAU91176	Pantothenate kinas
23	128.5	11.0	265	23	AAU91151	Streptomyces coeli
24	125.5	10.7	260	23	AAU91173	Pantothenate kinas
25	123.5	10.6	250	23	AAU91150	Treponema pallidum
26	122.5	10.5	273	23	AAU91157	Pantothenate kinas
27	120.5	10.3	212	23	AAU91177	Pantothenate kinas
28	117	10.0	233	23	AAU91163	Pantothenate kinas
29	114	9.8	272	22	AG81225	Mycobacterium tube
30	114	9.8	272	23	AAU91152	Mycobacterium tube
31	111.5	9.5	389	21	AAU74909	Neisseria meningit
32	108	9.2	455	21	AAU74910	Neisseria meningit
33	108	9.2	592	20	AAU38616	Neisseria meningit
34	108	9.2	592	21	AAU74913	Neisseria meningit
35	108	9.2	592	23	AAU91166	Pantothenate kinas
36	103	8.8	592	20	AAU38615	Neisseria meningit
37	103	8.8	592	21	AAU74912	Neisseria meningit
38	103	8.8	592	23	AAU91169	Pantothenate kinas
39	99	8.5	455	20	AAU38617	Neisseria gonorrh
40	99	8.5	455	21	AAU74908	Neisseria gonorrh
41	99	8.5	460	23	AAU91167	Pantothenate kinas
42	99	8.5	592	20	AAU38618	Pantothenate kinas
43	99	8.5	592	21	AAU74911	Neisseria gonorrh
44	98.5	8.4	248	23	AAU91164	Pantothenate kinas
45	97.5	8.3	484	18	AAW55720	H. pylori ORF 07ee

#### ALIGNMENTS

#### RESULT 1

ID AAU91159 standard; Protein; 229 AA.  
AC AAU91159;  
DT 05-JUN-2002 (first entry)  
DE Aquifex aeolicus pantothenate kinase Coax.  
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Aquifex aeolicus.  
XX WO200216601-A2.  
XX 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
XX 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -



Claim 10; Page 74-75; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence	246 AA;
SQ	

```

Query Match      16.68; Score 193.5; DB 23; Length 246;
Best Local Similarity 26.5%; Pred. No. 1.8e-11;
Matches 66; Conservative 45; Mismatches 105; Indels 33; Gaps 6;

3 FLTVDVGNSSVDIALMEGKKVKDFLKLSSHEEFKKE-----PPLKALGI-- 47
  : |||||: : : : || : : : : : : : : : : : : : : : : : :
2 YLLVDVGNTHSVESITDGTFRWRRLSTGVFOTDEFLSHLPLLGDMREIKGIVAS 61
  : |||||: : : : || : : : : : : : : : : : : : : : : : :

48 -----SVKQSFSEKVRGKTPKIFLKLKENFPQVDYKKTETLGTDRVALAYSAKKPYGK 101
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :
62 VVPTQNTVIERFSQKTFHISP--IWKAKNGCKVKNVKNPSEVGADRVANNVAEVKPYGK 119
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :
102 NVVVISAGTALVLDLVEGFKGFTFLGGLKKILKLSLDAEGIPEFPFPEVEIFLGRST 161
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :
120 NGIIIDMGATTVDLVVNGSYEGAILPGFFMVHSLFRGTAKPLPVEVKPADPVPVKDGT 179
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :
162 RECVLGAGYRESETEFIKTLKLRKPKFKFVVITGEGKYFS---KFGIYDPLLVHVRG 218
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :
180 EENIRLGVNGSVYALEGIIGRIKEVY-GDLPVVLGTGQSKIVKDMIKHIEFDEDLTJGK 238
  : : ||||: : : : : : : : : : : : : : : : : : : : : : :

219 MRNLLYLYH 227
  : : ||
239 -----VYH 241
  : : ||

```

RESULT 4  
AAU91165  
ID AAU91165 standard; Protein; 209 AA.

DT 05-JUN-2002 (first entry)

DE Pantothenate kinase (Coax) #3.

...  
KW Pantothenate kinase; Coax; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

OS Campylobacter jejuni.

AA  
PN WO200216601-A2.

AA PD 28-FEB-2002.

XX  
PF 24-AUG-2001: 2001WO-US26531.XX  
PR 24-AUG-2000: 2000US-227860P-

PR 20-MAR-2001; 2001US-0813453.  
XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
YY

PI Yocum RR, Patterson TA;

DR WPI; 2002-269358/31.

PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (CoaX) protein with test compound and identifying inhibitor of the CoaX protein -

XX  
PS Claim 6; Page 84-85; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (coax) protein described in the invention.

Sequence 209 AA;

Query Match	15.2%	Score 177.5;	DB 23;	Length 209;
Best Local Similarity	25.8%;	Pred. No. 6.3e-10;		
Matches 62;	Conservative 44;	Mismatches 83;	Indels 51;	Gaps 10;
Qy	4	LTVDVGNSSVDIALWEKQ----	KVKDFLKLSHSEFLKEEFPKPKALGISVKSQSESEKVRG	59
Db	2	LLCDIGNSNANF--LDDNKYFTLNIDQLEFKNEQ-----	KIFYINVNHLKHLK--	50
Qy	60	KIPKIKFLKKNPFIQVDYKTPET-----	LGTDVVALAYSAKKFKGKNVVVISAGTALVID	115
Db	51	-----NOKNFINLEPYEFLDTIYQGLGIDRIAACYTI-----	EDGVVVDAGSATIID	97
Qy	116	LVLEGPKFGFYTGLGKKLILSDLAEGIPPEFPEVEFI--	FLGRSTRECVLGGAYREST	174
Db	98	IIISNSIHLGFIILPGIANYKKIYSHISPRUKSENTQVSLDAPFKQTMDSALSYGFKGIY	157	
Qy	175	EFIKSTLKLWRKVKRKKFVVITGGEGKYSKFI--	GIYDPLLVHRGMR-----	NLLY 224
Db	158	LLIKDA-----	AQNKKIYFGGQGFILANYFDHAIYDKLLIFRGKKIKIENPNLLY	209

RESULT 5  
AAU91170  
ID AAU91170 standard; Protein; 262 AA.

AA  
AC  
AAU91170;

DT 05-JUN-2002 (first entry)

DE Pantothenate kinase (CoaX) #8.

XX  
pantothenate kinase: CoA: antibiologic; antimicrobial; XX

XX  
XX  
MW  
pantotenate kinase

XX  
XX  
XX

PN XX  
WU20021601-

XX  
FD 2007-2007.

XX  
FF  
Z4-AUG-2001; Z001WO-0520331.

PR 24-AUG-2000; 2000US-227860P;  
PR 20-MAR-2001; 2001US-0813453;

XX  
PA (OMNT-) OMNIGENE BIOPRODUCTS INC.

XX  
DT  
VOCUM PP  
Datterson TA.

XX  
DB 2002-260259/21

DR N-PSDB; ABK54191.

PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothene kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 8; Page 98-99; 128pp; English.

PS Claim 8; Page 98-99; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 XX Sequence 262 AA;

Query Match 14.3%; Score 167.5; DB 23; Length 262;  
 Best Local Similarity 23.7%; Pred. No. 9.2e-09;  
 Matches 61; Conservative 50; Mismatches 103; Indels 43; Gaps 9;  
 QY 6 VDVGNSVDIALWEGKKVXDFLKLSHEEFLKEEFPKLGKALGIVKQ-----SFSEKVR 58  
 b 5 LDVGNTNAVIGVPEEGELQOHMETDRHKTED-----EYGMVLVKOLLEHGLSF-EDVK 58  
 QY 59 G-----KIPKIF-----LAKENPPIQ-----VDYKTPETLGTDRVALAYS 96  
 Db 59 GIIVSSVVPIMPALERMCKEYKIKPLVVGPGIKTGLNLYKENPREVGADRVNAVAGI 118  
 QY 97 KFYGNVVISAGTALVIDLVLEGG-FKGGFITLGLGKKLKILSDLAEGIPEFPPEVEI 155  
 Db 119 HLYGSLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETIKPSS 178  
 QY 156 FLGRSTRECVLGAYRESTEFIKSTLKLWRKFKKFKVITGEGKYFSK-----FGIYD 211  
 Db 179 VGNKNTVAMQSGILYGVGOVEGIVKRMKEAKQEPKVIATGGLAKLISESNVIDVDV 238  
 QY 212 PLLVHRGMRLNLYLYHR 228  
 Db 239 PFLTLKG---LYMLYER 252

RESULT 6  
 AAU91179  
 ID AAU91179 standard; Protein; 241 AA.  
 AC AAU91179;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #17.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Thiobacillus ferrooxidans.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 XX 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX N-PSDB; ABK54200.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -

XX Claim 10; Page 109-110; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX

SQ Sequence 241 AA;  
 Query Match 14.1%; Score 164.5; DB 23; Length 241;  
 Best Local Similarity 29.5%; Pred. No. 1.7e-08;  
 Matches 46; Conservative 31; Mismatches 70; Indels 9; Gaps 4;  
 QY 77 DYKTPETLGTDR-RVALAYSARKFYGNVVISAGTALVIDLVLEGGFKGGFITLGLGKKL 135  
 Db 90 DYHPPESLGFDRRCCLLAAAMDYPGQDSIVIDMCTAITDILAGGHFRGRILPCIANSL 149  
 QY 136 KILSDLAEGIPEFPPEVEIFLGRSTRECVLGAYRESTEFIKSTLKLWRKFKKFKRVV 195  
 Db 150 RGLHECTALLPEVVLNAPAEMLGNDTSNAIQAGVHILFADALRGAITDFRQ-YSQARIL 208  
 QY 196 ITGEGEYFS---KFGIYDPLLVRHGRNLLYLYHR 228  
 Db 209 ITGDAERWOPGAGSLYQPHLLRG----FYLWIR 240

RESULT 7  
 AAU91153  
 ID AAU91153 standard; Protein; 258 AA.  
 AC AAU91153;  
 XX  
 XX 05-JUN-2002 (first entry)  
 XX Rhodobacter capsulatus pantothenate kinase Coax.  
 DE Rhodobacter capsulatus pantothenate kinase Coax.  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Rhodobacter capsulatus.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 XX 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 XX WPI; 2002-269358/31.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 10; Page 71-72; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an



CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.

XX Sequence 258 AA;

Query Match 13.6%; Score 158.5; DB 23; Length 258;  
 Best Local Similarity 22.0%; Pred. No. 7.5e-08;  
 Matches 57; Conservative 48; Mismatches 113; Indels 41; Gaps 9;  
 QY 4 LTVDVGNSSVDIALWEGKKVDF-----LKLSEEFLEKEP-----PKKALGIVSKQS 52  
 DB 3 LCIDCGNTWTFVSWDG---TDFAATWRIATDHRRTADEYFVWLNLMQLGQGRISEA 59  
 QY 53 FSEKVRGKIPKIKFLKE-----NFIQVDYKTPETLGTDRVALAYS 94  
 60 I---ISSVAPRVFNLVLCNRYFDCRPVYVGVKPCCELPVAPRVDPTGVGDLRVNTVA 116  
 QY 95 AKKFGKNNVISAGTALVIDLV-EGKFKGGFITLGLGKKILKILSDLAEGIPPEPPEV 153  
 DB 117 GYDRHGGDLIVVDFGTATTFVDPADGAYIGGVAPGVNLSLEALHMAAALPHVDVTKP 176  
 QY 154 EIFLGRSTRCVLGGAYRESTEFIKSTLKLWRKVKFKKVVITGEGKYFS-KFGIYDP 212  
 DB 177 QGVIGTNTVACIQSGVYWGVIYGLVEGIVRQIMRDRPMKVIATGGLASLFDLGFDFDK 236  
 QY 213 L---LVHRGMRLNLYLHR 228  
 DB 237 VEDDLTMHGLR-LIFDYNK 254

# RESULT 8

ABB47661  
 ID ABB47661 standard; Protein; 259 AA.

AC ABB47661;

XX 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #365.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP ) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusnick C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides

XX Claim 6; SEQ ID NO 366; 192pp; French.

XX The present invention relates to the genome sequence of *Listeria*  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by *L.*  
 CC monocytogenes and related organisms.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 259 AA;

Query Match 13.4%; Score 156.5; DB 23; Length 259;  
 Best Local Similarity 23.8%; Pred. No. 1.2e-07;  
 Matches 61; Conservative 39; Mismatches 113; Indels 43; Gaps 7;

QY 4 LTVDVGNSSVDIALWEGKKVDFLKLSEEFLEKEPFPKALGIVSKQSFSEK----- 56  
 DB 3 LVIDVGNTCTGVYEKQL-----LKHWRMTTDRHRTSDELGMVTLNFFSYANLTFSDI 57  
 QY 57 -----VRGKIPKIKFLK-----KENFFIQVDYKTPETLGTDRVALAY 93  
 DB 58 QGIILSSVVPPIHMAETMCVRYFNIRPLIVGPGIKTGLNLKVD--NPREIGSDRIYNAV 115  
 QY 94 SAKKFGKNNVISAGTALVIDLVLP-GKFKGGFITLGLGKKILKILSDLAEGIPPEPPEE 152  
 DB 116 AASEYGTPIVVDFGTATTFYIDESGVYGGAIAPGIMISTEALYNRAKLPVRDIAE 175  
 QY 153 VEIFLGRSTRCVLGGAYRESTEFIKSTLKLWRKVKFKKVVITGEGKYF-----SKFG 208  
 DB 176 SSQIIGKSVSSMQAGIFYGVGQCCEGIIAEMKKQSNASPVVATGGLARMIKSSAVD 235  
 QY 209 IYDPLLVHRGMRLNLY 224  
 DB 236 ILDPFLTKGL-ELLY 250

## RESULT 9

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX AAU01243;

XX 18-JUL-2001 (first entry)

XX B. subtilis novel pantothenate kinase encoded by the gene coax.

XX Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
 KW nutritional supplement; panto-compound; pantoate.

OS Bacillus subtilis.

XX WO200121772-A2.

XX 29-MAR-2001.

XX 21-SEP-2000; 2000WO-US25993.

XX 21-SEP-1999; 99US-0400494.

XX 07-JUN-2000; 2000US-0210072.

XX 28-JUL-2000; 2000US-0221836.



```
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR N-PSDB; ABK54193.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein
XX S Claim 10; Page 101-102; 128pp; English.
XX S
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 258 AA;
Query Match 12.7%; Score 148.5; DB 23; Length 258;
Best Local Similarity 23.9%; Pred. No. 7.9e-07;
Matches 60; Conservative 43; Mismatches 111; Indels 37; Gaps 9;
QY 6 VDVGNSVDIALWEGKKVDF--LKLSEEFLEKEFPKALKGIVSKQSFSEKVRG----- 59
Db 5 LDVGNTVGLVYDGDGLKHHWRIETSRKTEDEYGMIMKALNLHVGLQFSD-IRGIIS 63
QY 60 -KIPKIKFLKE-----NFIQVDYKPTLGTDRVALAYSARKFYK 101
Db 64 SVVPPFMALERMCLYFHKPLIVGPGIKTGIDIKIDNPREGVAGDRIVNAVAGIHLYGS 123
QY 102 NVVISAGTALVDLVLEGG-FKGGFTITLGLGKKLILSDLAEGIPPEFFPEVEIFLGRS 160
Db 124 PLIIVDFGATTVCYINEHKQYMGGAIAFGIMISTEALFARAACLPRIEIARPDIIIGN 183
QY 161 TRECVLGG---AYRETEPIKSTLKLWRKVKRKKVITGGEGKYFSK-----FGYDPL 213
Db 184 TVSAMOAGILYGVGVEGIVSRMAKSKI---PPKVIATGGIAPLIASESDIIDVDVDF 240
QY 214 LVHGRMRLLY 224
Db 241 LTLTGLK-LLY 250
RESULT 12
ID AAU91158 standard; Protein; 262 AA.
XX AC AAU91158;
XX DT 05-JUN-2002 (first entry)
XX DE Borrelia burgdorferi pantothenate kinase Coax.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Borrelia burgdorferi.
```

```
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX DR WPI; 2002-269358/31.
XX DR
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein
XX S Claim 6; Page 76; 128pp; English.
XX S
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 262 AA;
Query Match 12.1%; Score 141.5; DB 23; Length 262;
Best Local Similarity 28.3%; Pred. No. 4.2e-06;
Matches 62; Conservative 38; Mismatches 94; Indels 25; Gaps 9;
QY 4 LTVDVGNSSVDIALWEGKKVDFLK-----LSHEE---FLKE--EPPKALKGIVSKQS 52
Db 9 LIIDIGNTSIAFLKDNQVNLFIKMTNLMRLYDEVYSFEEFNENKVFISVVPI 68
QY 53 FSEKVRGKIPKIFLKE---NFIQVD-----YKTPF-TLGTDRVALAYSARKFYK-K 101
Db 69 LNETFRNVI--FSFFKIKPLFIFGLDNLFTFNPKSKDFLLGSDVFNALVAIENYSFE 126
QY 102 NVVISAGTALVDLV-LEGKFGGFTITLGLGKKLILSDLAEGIPPEFFPEVEIFLGRS 160
Db 127 NVLVDLGTACTFVRSRQDGIIGIINSGLINFNLSLLDNAYLIKFPPISTPNLLERT 186
QY 161 TRECVLGGAYRETEPIKSTLKLWRKVKRKKVITGG 199
Db 187 TGSVNSGLFYQYKYLIEGVYRDIKQYKKKNLIITGG 225
RESULT 13
ID AAU91154 standard; Protein; 255 AA.
XX AC AAU91154;
XX DT 05-JUN-2002 (first entry)
XX DE Geobacter sulfurreducens pantothenate kinase Coax.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Geobacter sulfurreducens.
XX PN WO200216601-A2.
XX
```

PD XX 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX XX WPI; 2002-269358/31.

DR XX Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein -

XX XX

PS Claim 10; Page 72-73; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX XX

SQ Sequence 255 AA;

Query Match 11.9%; Score 139.5; DB 23; Length 255;

Best Local Similarity 22.1%; Pred. No. 6.5e-06;

Matches 57; Conservative 47; Mismatches 113; Indels 41; Gaps 9;

QY 4 LTVDGNSVDIALWECKKVKDFLKLISHEE-----FLKEEPPKALKGIS-----VK 50

DB 3 LVIDVGNNTVLGIYDGERLVDRMVSTDKARTTDEYGLINELFRLAGLGLQIRAVII 62

QY 51 QSFSEKVRGKIPKIF-----LK-----KENFPIQVDYKTPETLGTDRVALAYSAKKF 98

DB 63 SSVVPELTGVLRLSLGFGMRPLVGVPGIKTGMPIQ--YDNREVGAADRVINAVAGYEK 120

QY 99 YGNVNVISAGTALVIDLV-LEGKFKGGFTTLGLGKKLKLISDLAEGIPEFFPEVEIFL 157

DB 121 YRTSLIIVDGTATTEDYVNRKGEYCGAIAAGLVISTEALFORASKLPRVDIIRPSAII 180

QY 158 GRSTRECVLGGAYRESTEFIKSTLKLWRKVKFKFKVVIWG-----EGKVFSEKGIY 210

DB 181 ARNTVNSMQAGIYYGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTIIEAVEEY 240

QY 211 DPLLVHRGMNLLYLHR 228

DB 241 ---LTLEGLR---ILYER 252

RESULT 14

AAW98422

ID AAW98422 standard; Protein; 223 AA.

XX AC AAW98422;

XX 31-MAR-1999 (first entry)

XX H. pylori GPHO 344 protein.

XX GPHO protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14141.

XX New isolated Helicobacter polynucleotides - used to develop products

PT for the diagnosis, prevention and treatment of Helicobacter

PT infections and gastrointestinal diseases

XX XX

PS Claim 8; Page 709-710; 2054pp; English.

XX This sequence represents a Helicobacter pylori GPHO protein of the

CC invention. The polypeptides can be used for preventing or treating

CC Helicobacter infections, and gastroduodenal diseases associated with

CC these infections, including acute, chronic, and atrophic gastritis, and

CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be

CC used for the production of antibodies. The products can also be used for

CC detection and diagnosis.

XX XX

SQ Sequence 223 AA;

Query Match 11.9%; Score 138.5; DB 19; Length 223;

Best Local Similarity 26.2%; Pred. No. 6.8e-06;

Matches 59; Conservative 35; Mismatches 88; Indels 43; Gaps 10;

QY 7 DVGNSVDIALWECKKVKDFLKLISHEEFLKEEPPKALKG-----ISVKQSFSEKVRG 59

DB 17 DIGNTRIHEA-----QNYQLFSSAKEDLRLGIQKEIFYISVNEENEKALLN 63

QY 60 KIPKIKELKENPFIQVDYKTPETLGTDRVALAYSAKKFYKNVVISAGTALVIDLVLE 119

DB 64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVDVAGSAITDLIKE 114

QY 120 GKFKGGFTTLGLGKKLKLISDLAEGIPEFFP--EVEIFLGRSTRECVLGGAYRESTEF 177

DB 115 GKHLGGCILPLAGIYHAYKKSAILKIQPFKALDSLEV-LPKSTRDVNYG-----MV 166

QY 178 KSTLKLWRKVKFKFKVVIWGGEKGYFSKF---GIYDPLLVHRGM 219

DB 167 LSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210

RESULT 15

AAU91161

ID AAU91161 standard; Protein; 223 AA.

XX AC AAU91161;

XX 05-JUN-2002 (first entry)

XX Helicobacter pylori pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Helicobacter pylori.

XX WO200216601-A2.

XX 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.  
XX  
XX 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
XX Yocum RR, Patterson TA;  
XX  
XX WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 6; Page 78-79; 128pp; English.  
XX  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 223 AA;  
Query Match 11.9%; Score 138.5; DB 23; Length 223;  
Best Local Similarity 26.2%; Pred. No. 6.8e-06;  
Matches 59; Conservative 35; Mismatches 88; Indels 43; Gaps 10;  
QY 7 DVGNSVDIALWEGKKVDFLKLSEEFLEKKEFPKALG-----ISVKQSFSEKVRG 59  
Db 17 DIGNTRIHA-----QNYQLFSSAKEDLKLGIQKEIFYISVNEENEKALLN 63  
QY 60 KIPKIKFLKENFPIQVDYKTPETLGTDRVALAYSAKFYGNVNVVISAGTALVIDLVLE 119  
Db 64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITDLIKE 114  
QY 120 GKFKGGFTLGLGKKLKLSDLAGIPEFFP--EVEIFLGRSTRECVLGGAYRESTEFI 177  
Db 115 GKHLGGCFLPLGQYIHAYKKSALILEQPFKALDSLEV-LPKSTRDAYNYG-----MV 166  
QY 178 KSTLKLWRKFKRKFKVITGEGGKYSKF---GIYDPLLVHRGM 219  
Db 167 LSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVDGM 210

Search completed: June 24, 2003, 21:46:20  
Job time : 24.511 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 13.3935 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-63  
Perfect score: 1264  
Sequence: 1 MIFIAVGNTRILLATHDGV.....AGSLYQPHLLRGFLYINRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

earched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1264	100.0	241	9 US-09-813-453A-63	Sequence 63, Appl
2	229.5	18.2	246	9 US-09-813-453A-9	Sequence 9, Appl
3	227	18.0	257	9 US-09-813-453A-53	Sequence 53, Appl
4	227	18.0	260	9 US-09-813-453A-51	Sequence 51, Appl
5	215	17.0	262	9 US-09-813-453A-8	Sequence 8, Appl
6	203	16.1	258	9 US-09-813-453A-2	Sequence 2, Appl
7	190.5	15.1	262	9 US-09-813-453A-45	Sequence 45, Appl
8	189	15.0	219	9 US-09-813-453A-57	Sequence 57, Appl
9	184	14.6	265	9 US-09-813-453A-4	Sequence 4, Appl
10	183.5	14.5	244	9 US-09-813-453A-41	Sequence 41, Appl
11	181	14.3	233	9 US-09-813-453A-17	Sequence 17, Appl
12	180.5	14.3	212	9 US-09-813-453A-59	Sequence 59, Appl
13	179.5	14.2	254	9 US-09-813-453A-47	Sequence 47, Appl
14	178.5	14.1	258	9 US-09-813-453A-49	Sequence 49, Appl
15	174	13.8	257	9 US-09-813-453A-13	Sequence 13, Appl
16	172.5	13.6	255	9 US-09-813-453A-7	Sequence 7, Appl
17	170.5	13.5	273	9 US-09-813-453A-10	Sequence 10, Appl
18	169	13.4	250	9 US-09-813-453A-3	Sequence 3, Appl
19	168.5	13.3	258	9 US-09-813-453A-6	Sequence 6, Appl

20	164.5	13.0	229	9	US-09-813-453A-12	Sequence 12, Appl
21	159	12.6	272	9	US-09-712-363-276	Sequence 276, Appl
22	159	12.6	272	9	US-09-813-453A-5	Sequence 5, Appl
23	154.5	12.2	267	9	US-09-813-453A-15	Sequence 15, Appl
24	145	11.5	256	9	US-09-813-453A-55	Sequence 55, Appl
25	144.5	11.4	223	9	US-09-895-913A-74	Sequence 74, Appl
26	144.5	11.4	223	9	US-09-813-453A-14	Sequence 14, Appl
27	144.5	11.4	223	9	US-09-813-453A-67	Sequence 67, Appl
28	144.5	11.4	248	9	US-09-813-453A-20	Sequence 20, Appl
29	143.5	11.4	242	9	US-09-813-453A-65	Sequence 65, Appl
30	136	10.8	262	9	US-09-813-453A-11	Sequence 11, Appl
31	124	9.8	209	9	US-09-813-453A-21	Sequence 21, Appl
32	123.5	9.8	460	9	US-09-813-453A-39	Sequence 39, Appl
33	115.5	9.1	592	9	US-09-813-453A-43	Sequence 43, Appl
34	112.5	8.9	592	9	US-09-813-453A-22	Sequence 22, Appl
35	107.5	8.5	249	9	US-09-813-453A-70	Sequence 70, Appl
36	104.5	8.3	249	9	US-09-813-453A-61	Sequence 61, Appl
37	94.5	7.5	1827	9	US-09-712-363-261	Sequence 261, Appl
38	94.5	7.5	5215	9	US-09-860-846-2	Sequence 2, Appl
39	94.5	7.5	5215	9	US-09-988-384B-2	Sequence 2, Appl
40	94.5	7.5	5215	9	US-09-836-821-2	Sequence 2, Appl
41	94.5	7.5	5215	10	US-09-861-289-2	Sequence 2, Appl
42	89	7.0	507	9	US-10-270-333-45	Sequence 45, Appl
43	86	6.8	732	9	US-10-138-927-78	Sequence 78, Appl
44	86	6.8	1077	9	US-10-121-911-1	Sequence 1, Appl
45	85	6.7	760	9	US-09-712-363-292	Sequence 292, Appl

## ALIGNMENTS

## RESULT 1

US-09-813-453A-63  
Sequence 63, Application US/09813453A  
Patent No. US20020168681A1

GENERAL INFORMATION:  
APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.  
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

TITLE OF INVENTION: ANTIBIOTICS  
FILE REFERENCE: OGZ-001

CURRENT APPLICATION NUMBER: US/09/813,453A  
CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: US 60/227,860  
PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569  
PRIOR FILING DATE: 2000-09-21

NUMBER OF SEQ ID NOS: 77  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 63  
LENGTH: 241

TYPE: PFT  
ORGANISM: Thiobacillus ferrooxidans

US-09-813-453A-63

Query Match 100.0%; Score 1264; DB 9; Length 241;  
Best Local Similarity 100.0%; Pred. No. 6.1e-113;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MIFIAVGNTRILLATHDGVHSDVSVA	1	MIFIAVGNTRILLATHDGVHSDVSVA	1	Sequence 63, Appl
DB	1	MIFIAVGNTRILLATHDGVHSDVSVA	1	MIFIAVGNTRILLATHDGVHSDVSVA	1	Sequence 9, Appl
QY	61	ALAARREALATAEVEPDGFFRRVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	61	ALAARREALATAEVEPDGFFRRVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	61	Sequence 53, Appl
DB	61	ALAARREALATAEVEPDGFFRRVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	61	ALAARREALATAEVEPDGFFRRVPHDYHPPESLGFDRRCCLLAAMDYPGQDSIVI	61	Sequence 51, Appl
QY	121	DMGTAITIDLLAGGHFRGGRILLPGIAMSRLGHEGTALLPEVVLNAPAEMLGNDTSNAIQ	121	DMGTAITIDLLAGGHFRGGRILLPGIAMSRLGHEGTALLPEVVLNAPAEMLGNDTSNAIQ	121	Sequence 8, Appl
DB	121	DMGTAITIDLLAGGHFRGGRILLPGIAMSRLGHEGTALLPEVVLNAPAEMLGNDTSNAIQ	121	DMGTAITIDLLAGGHFRGGRILLPGIAMSRLGHEGTALLPEVVLNAPAEMLGNDTSNAIQ	121	Sequence 2, Appl
QY	181	AGVHLFDALRGATDFDFOYSPQARILITGGDAERQPGIAGSLYQPHLLRGFLYINRW	181	AGVHLFDALRGATDFDFOYSPQARILITGGDAERQPGIAGSLYQPHLLRGFLYINRW	181	Sequence 45, Appl
DB	181	AGVHLFDALRGATDFDFOYSPQARILITGGDAERQPGIAGSLYQPHLLRGFLYINRW	181	AGVHLFDALRGATDFDFOYSPQARILITGGDAERQPGIAGSLYQPHLLRGFLYINRW	181	Sequence 57, Appl

```
Db 181 AGVHFLFADALRGAITDFRQYSPQARILITGDAERWQPGIAGSLYQPHLLRGFYLR 240
QY 241 G 241
Db 241 G 241

RESULT 2
US-09-813-453A-9
; Sequence 9, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-09-813-453A-9
```

```
Query Match 18.2%; Score 229.5; DB 9; Length 246;
Best Local Similarity 28.2%; Pred. No. 4e-14;
Matches 71; Conservative 40; Mismatches 110; Indels 31; Gaps 5;

QY 2 IFIAGNTRTLAHTHDGVHFDVSVATSLPTEILQOQGL-TWLSAPNREPVALG--GV 58
Db 3 LLYDVGNTHSVETEDGKTFRRWLSTGVFTDELESHLPLLDGAMREIKGICVASV 62

QY 59 VPAALAA-----WREALATAEVREPDPGFFRRVAVPHDYHPPESLGFDRRC 104
Db 63 VPTQNTVIERFSQYFHSPIWVAK-----NGCVKNVKNPSEVGADRVAN 109

QY 105 LLAAMDYPGQDSIVIDMGTAITIDLLAGHFRGGRILPGIAMSRLGLHEGTALLPEVVL 164
Db 110 VVAFVKEY-GKNGIIDMGTAITVDLVNNGSYEGGAILPGFFMVMVHSLFRGTAKPLVEV 168

QY 165 NAPAEMLGNDTSNAIQAGVHFLFADALRGAITDFRQYSPQARILITGDAERWQPGIAGS 224
Db 169 KPADFVVGKDTENIRLGVNNGSVYALEGIIRIKEVYGDLPVLTGGOSKIVKDMIKHE 228

Y 225 LYQPHLLRGFY 236
Db 229 IFDEDITKGVY 240
```

```
RESULT 3
US-09-813-453A-53
; Sequence 53, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Chlorobium tepidum
US-09-813-453A-53

Query Match 18.0%; Score 227; DB 9; Length 257;
Best Local Similarity 32.1%; Pred. No. 7.3e-14;
Matches 84; Conservative 43; Mismatches 101; Indels 34; Gaps 12;

QY 2 IFIAGNTRTLAHTHDGVHFDVSVATSLPTEILQOQGLTWLSAPNREP--VALGGV 58
Db 3 LLYDVGNTHSVETEDGKTFRRWLSTGVFTDELESHLPLLDGAMREIKGICVASV 61

QY 59 VPAALAA-----WREALATAEVREPDPGFFRRVAVPHDYHPPESLGFDRRCCLLAA 108
Db 62 VPSATAVGSALESLESFVPLTICCKLRFP-----FRL---DYATPHTFGADRLALCAWS 113

QY 109 AMDYPGQDSIVIDMGTAITIDLL-AGHFRGGRILPGIAMSRLGLHEGTALLPEVVLNAP 167
Db 114 RHLFSEKPVIAVDIGTAITFDVLDTVGNRYRGLIMPGIDMMAGALHSRTAQLPOVRIDRP 173

QY 168 AEMLGNDTSNAIQAGV-----IHLFADALRG-AITDFRQYSPQARILITGDAERWQ 219
Db 174 ESSLGRSTTECKSGVFWGVYQIGGLVDAIRGDLVNDFGEST--VEVITGGNSRIIVP 231

QY 220 GIAG-SLYQPHLLRGFYLR 240
Db 232 EIGPVSVIDEVLAVLRGSDLLLR 253
```

```
RESULT 4
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-813-453A-51
```

```
Query Match 18.0%; Score 227; DB 9; Length 260;
Best Local Similarity 29.0%; Pred. No. 7.4e-14;
Matches 75; Conservative 43; Mismatches 103; Indels 38; Gaps 12;

QY 1 MIFIAGNTRTLAHTHDG-----VHFDVSVATSLPP-----TEILQOQGLTWLSAPNR 50
Db 2 LLAIEQGNNTNFA-IHDGASVWAQWSATSTRTADEYVWVLSQLLSMOGLGFRAI--- 57

QY 51 EPVALGGVYVPAALAAWR-----EALATAEVREPDPGFFRRVAVPHDYHPPESLGFDR 101
Db 58 DAVIISVVYQSQIFNLRLNRRYFNVEPLVIGENA-----KLGDVRIEKPSEAGADR 110

QY 102 RCCLLAAMDYPGQDSIVIDMGTAITIDLLAG-GHFRGGRILPGIAMSRLGLHEGTALLP 160
Db 111 LVNAIGCAAMVYFG-PLVVIDSGTATTFDIVAADCAEGGIIAAGINLSMALHEAAKLP 169

QY 161 EYVLNAPA--EMLGNDTSNAIQAGVHFLFADALRGAITDFR-QYSPQARILITGDAERW 217
```



Db 170 RIAORPAGNRIVGTDTSAMQSGVFWGISLIEGLVARIKAEGERPMTVIATGGVASLF 229  
QY 218 QPGIAGSL--YOPHLLLRG 234  
Db 230 E-GATDSIDHFDSDLITNG 247

## RESULT 5

US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8

Query Match 17.0%; Score 215; DB 9; Length 262;

Best Local Similarity 29.7%; Pred. No. 1.1e-12;  
Matches 78; Conservative 35; Mismatches 112; Indels 38; Gaps 11;

QY 1 MIFIAVGNTRTLL-----ATHDGVHDSVSAT-----LPTTELQOPGL-TWLSAPN 49  
Db 6 LIAVDIGNTTVLGLADASGALH-----TWIRTNREMLPDDLALQLHGLFTLAGAPI 59  
QY 50 REPVALGGVPA-----ALAAWR-----EALATAEVRPDPGFFRRAVPHDYHPPESLGF 100  
Db 60 PRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPD-----VTVELDTPGSGVAD 112  
QY 101 RCCLLAAMDPGQD-STIVDMGTAITIDLLA-GGHFRGGRILPGIAMSRLRGLHGTAL 158  
Db 113 RLCNLFGBKYLGLLDYAVVDFGTSTNFDVVGRRRFLGGLILATGAQVSADALFARA 172  
QY 159 LPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQ 218  
Db 173 LPRITLQAPETAIGNTKNTVALOSGLVFGYAEWVDGLLRIRRAELPGEAVAVATGGPSRTV 232  
QY 219 QGIAGSL--YOPHLLLRGFY-LW 238  
Db 233 QGICQEIYDYDTLRLGLVELW 255

## RESULT 6

US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2

Query Match 16.1%; Score 203; DB 9; Length 258;

Best Local Similarity 28.3%; Pred. No. 1.4e-11;  
Matches 73; Conservative 38; Mismatches 115; Indels 32; Gaps 10;

QY 1 MIFIAVGNTRTLLAHTHDGVHDSVSATSLPTE-----ILQOPGLTWLSAPNR 51  
Db 2 LVIDVGNNTVLGVYHDGKLEYHWRIETSRHKTDEFGMLRSLFPHSGLMF-----EQID 58  
QY 52 PVALGGVPAALAAWREALATAEVRPDP-----PGFFRRAVPHDYHPPESLGRDRCCLLA 108  
Db 59 GIIISVVPPIMFALERMCKYFHEIPQIVGPG-MKTGLNIKYDNPKREVGAADRVNAVA 117  
QY 109 AMDYPGQDSIVDMGTAIT-----IDLLAGHFRGGRILPGIAMSRLRGLHGTALLPEVVLN 165  
Db 118 IHLV-GNPLIVDFGTATTYCYID--ENKQYMGGAIAPIGTISTEALYSRAAKLPRIET 174  
QY 166 APAEMLGNDTSNAIQAGVIHLFADALRGAITDFR-QYSPQARILITGGDAERWOPGIAG- 223  
Db 175 RPDNIIGKNTVSAMQSGILFGYGVQVEGIVRKMKAQKQDLKVIATGGLA-----PLIANE 230  
QY 224 ----SLYQPHLLLRGFY 237  
Db 231 SDCIDIVDPFLTKGLEL 248

## RESULT 7

US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 15.1%; Score 190.5; DB 9; Length 262;

Best Local Similarity 26.1%; Pred. No. 2.3e-10;  
Matches 66; Conservative 48; Mismatches 118; Indels 21; Gaps 9;

QY 1 MIFIA-VGNTRTLLAHTHDGVHDSVSATSLPTE-----ILQOPGLTWLSAPNR 50  
Db 1 MIFVLDVGNNTVLGVFEEGELRQHRMETDRHKTDEYGLMVKOLLEH-----LSFEDV 57  
QY 51 EPVALGGVPAALAAWREALATAEVRPDP-----PGFFRRAVPHDYHPPESLGRDRCCLLA 107  
Db 58 KGIIVSSVPPIMFALERMCKYKIKPLVVGPG-ITGLNIKYNPENREVGAADRVNAVA 116  
QY 108 AAMDYPGQDSIVDMGTAITIDLL-AGGHFRGGRILPGIAMSRLRGLHGTALLPEVVLNA 166  
Db 117 GIHLV-GSPLIIVDFGTATTYCYINEEKHYNGVITPGIMISAEALYSRAAKLPRIETK 175









```

Db      3 LLVDVGNTHSVFSTEDGKTFRRWRLSTGVFTQDELFSLHLLGLDAMREIKIGVASV 62
QY      59 VPAALAA-----WREALATAEVREPDGFFRRVAPVDHYHPPESLGDRRC 104
Db      63 VPTQNTVIERFSQYFHSPIWVK-----NGCVKWNKPNSEVGDVRVAN 109
QY      105 LLAAMDPGQDSIVDMGTATIDLAGHFRGGRILPGIAMSRLGHEGTALLPEVVL 164
Db      110 VVAFVEY-GRNGIIDMTATVDLVVNGSYEGGAILPGFFMVHSLFRGTAKLPLEV 168
QY      165 NAPAEMLGNDTSNAIQAGVHILFADALRGAIIDFRQYSPQARILITGGDAERWQPGIAGS 224
Db      169 KPADFVVGKDTENIRLGVVNGSVYALEGIIGRIKEIVGDLFPVLITGGQSKIVKMKIHE 228
QY      225 LVQPHLLRGFY 236
Db      229 IPFEDLTIKGVY 240

RESULT 2
Q9A621 PRELIMINARY; PRT; 261 AA.
Q9A621;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smith J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -
DR TIGR; CC1935; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
QY Complete proteome.
QJ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 18.0%; Score 228; DB 16; Length 261;
Best Local Similarity 29.1%; Pred. No. 1.le-10;
Matches 76; Conservative 44; Mismatches 101; Indels 40; Gaps 13;

QY 1 MIFIAV--GNRTTLAHTHDG---VHFDVSVAATSLPP-----TEILQOPLTWLSAP 48
Db 1 MMLIAIEGNTWTFP-LHDGASVQAQRSAESTRTADEYVWLSQLLSMOGLGFRAT- 58
QY 49 NREPVALGGVVPAALAAWR-----EALATAEVREPDGFFRRVAPVDHYHPPESLG 99
Db 59 --DAVIISVVQSFNRLNLSRRYFNVEPLVIGENA-----KLIDVRIKFPSEGA 109
QY 100 DRCCLLAAMDYPQDSIVDMGTATIDLAG-CHFRGGRILPGIAMSRLGHEGTAL 158
Db 110 DRLVNAIGAAMVYPG-PLVWIDSGTATTFDVAADGAFEGGIAPGINLSMOALHEAAK 168
QY 159 LPEVVLNAPA--EMLGNDTSNAIQAGVHILFADALRGAIIDFRQYSPQARILITGGDAE 215
Db 169 LPRIAQRPNRIVGTDTVSAMQSGVFWGYSISLIEGLVARIKARGEPMTVIATGGVAS 228

```

```

QY      216 RWQPGIAGSL--YQPHLLRG 234
Db      229 LFE-GATSDIDHFDSDLITNG 248

RESULT 3
Q9RX54 PRELIMINARY; PRT; 262 AA.
Q9RX54;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001905; AAF10040.1; -
DR TIGR; DR0461; -
DR InterPro: IPR004619; Baf.
DR InterPro: IPR001220; Lectin_legB.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;

Query Match 17.0%; Score 215; DB 16; Length 262;
Best Local Similarity 29.7%; Pred. No. 1.2e-09;
Matches 78; Conservative 35; Mismatches 112; Indels 38; Gaps 11;

QY 1 MIFIAVGNTRTLL-----AHTHDGVHFDVSVAATSLPPTEILQOPLG-TWLSAPN 49
Db 6 LLAVDIGNTTVILGLADASGALTH-----TWIRTNREMLPDDLAQLHLGLFTLAGAPI 59
QY 50 REPVALGGVVPA-----ALAAWR-----EALATAEVREPDGFFRRVAPVDHYHPPESLG 100
Db 60 PRAAVLSSVAPPVGENVALAKRHEFMDAFVSAENLPD-----VTVELDTPGSVGAD 112
QY 101 RRCLLAAAMDYPQD-SIVDMGTATIDLLA-GHFRGGRILPGIAMSRLGHEGTAL 158
Db 113 RLCNLFGAELKGLLDYAVVDFGTSTNFDVVGRRFLGGILATGAQVSADALFAAAK 172
QY 159 LPEVVLNAPAEMLGNDTSNAIQAGVHILFADALRGAIIDFRQYSPQARILITGGDAERWQ 218
Db 173 LPRITLQAPETAIGKNTVHALOSGLVFGYAEWDLGLLRIRAEPLGAEVAVATGGFGRTV 232
QY 219 PGIAGSL--YQPHLLRGFY-LW 238
Db 233 QGICQCIDYDETILRLGLVELW 255

RESULT 4
Q8RFE4 PRELIMINARY; PRT; 256 AA.
Q8RFE4;
AC Q8RFE4;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

```

01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
-OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Bhattacharya A., Bartman A., Reznik G., Los T., Lykidis A.,  
RA Bhattacharya A., Kogan I., Gardner W., Grechkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan I., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fongstein M., Kyripides N., Overbeek R.,  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586".  
RL J. Bacteriol. 184:2005-2018(2002).  
R EMBL; AE010586; AAU94957.1; -  
W complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
Query Match 16.1%; Score 203; DB 16; Length 256;  
Best Local Similarity 25.5%; Pred. No. 1.1e-08;  
Matches 6; Conservative 44; Mismatches 110; Indels 42; Gaps 8;  
QY 1 MIFAVGNTRTLLA-HTHDGVHFDVSVSATSLPTE-----LQOPGLTWLSAPNREP 53  
DB 2 IIGDIGNTHVTGIYDNGELISTFRIATNDRKTEDEYFSYFNNTKYNEISIKKVD 61  
QY 54 ALGVVPAALANR-----BALATAEVREPDPGFFRAVPHDYHPHESLGFDRCR 104  
DB 62 LISSVVPNIITFOFFARKYKVEA-TIVDLKKLPFTFAKGINY-----TGFGADRI 115  
QY 105 LLAAMDYPGDSIVIDMGTAITDLAGHGFRGRTLPGLAMSLRGHSGTALLPEVVL 164  
DB 116 ITEAQYDPKLVIFDFGATYDVLUKGYIGGGILPGIDMSINALYNTAKLPKVF 175  
QY 165 NAPAEMLGNTSNAIQAGVHFLFADALRGATIDF-RQYSPQARILITGDAERWQPGIAG 223  
DB 176 TTPSSVLGTDMKQIAIFPGYAGQIKHIKKINEELNEIFVLATGG-----LG 226  
QY 224 SL-----YQPHLLRGFY 236  
DB 227 KILSAEIDEIDEXDANLSLGLY 249  
RESULT 5  
J8YQD7  
ID Q8YQD7 PRELIMINARY; PRT; 276 AA.  
AC Q8YQD7;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein Alr3896.  
GN ALR3896.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriquchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpso S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.,  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium Anabaena sp. strain PCC 7120".  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003594; BAB/5535.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.

Db 2 ILVIDGNTCTGVYVYKQKLLKHWMTTDR-HRTSDELGMV-----LNFFSYA 50  
 QY 49 NREP-----VALGGVVPVPAALAAWREALATAEVR-----EP---DPGFFRAVPHDYHPPES 96  
 Db 51 NLTPSDIOGIITSSVPPIM-----HAMETMCVRYFNIRPLIVPGP-IKTGLNKLKVDNPRE 105  
 QY 97 LGFDRCCILAAADYPGQDSIVDMGTAITIDLL-AGHFGRGRLPGIAMSRLGLHEG 155  
 Db 106 IGSDRVNVAASEEY-GTPVIVDFGTATTCYIDESVYGGGAIPGIMISTEALYNR 164  
 QY 156 TALLPEVYNLPAEMLGNDTSNAIQGVTHLFADALRGAITDFROYSPQARILITGGDAE 215  
 Db 165 AAKLPVDAESSQIIGKTSVSSMQAGIFYGVGQCEGIIAEMKKQSNASPVVATGGIA 224  
 QY 216 RW-OPGIAGSLYQPHLLRGFYLR 240  
 Db 225 RMITEKSAVDILDPFLTKGLELLYR 251

## RESULT 7

QYX8N6 PRELIMINARY; PRT; 265 AA.  
 QYX8N6;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Hypothetical protein SC03380.  
 GN SC03380 OR SCE94.31C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 J. Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL049628; CAB40880.1; --  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs; TIGR00671; baf; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 14.6%; Score 184; DB 16; Length 265;  
 Best Local Similarity 28.8%; Pred. No. 3.8e-07;  
 Matches 76; Conservative 39; Mismatches 113; Indels 36; Gaps 11;  
 QY 1 MIFIAVGNTRTLIAHPTHG-----VHFDSVSVATSLPPTTEILO-----QPGLTWLSA 47  
 Db 2 LITIDVGNTHVLG-LFDGEDIVEHWRISTDSRTADEL--AVLQGLMGMHPLLGDELG 58  
 QY 48 PNREPVALGGVVPVPAALAAWREALATAEVR-----DPGFFRAVPHDYHPPESLGFDRRC 103  
 Db 59 DGIDGIAICATCVSVLHELREVTTRYGVDPVAVLVEPG-VKTGVPIITDHPKEVGADRII 117  
 QY 104 CILAAADYPGQDSIVDMGTAITIDLL-AGHFGRGRLPGIAMSRLGLHEGTALLPEV 162  
 Db 118 NAVAARELV-GGPAIVYDFGTATTFDAVSARGEIYIGVIAPGIEISVEALGVKGAQRKI 176  
 QY 163 VLNAPAEMLGNDTSNAIQGVTHLFADALRGAIT-----DFROYSPQARILITGGDAERWQ 218  
 Db 177 EVARPSVIGKNTVEAMQSGIYVGFAGQVDGVVNRVARELADDPDVTATGGLA----- 232  
 QY 219 PGIAGSL-----YQPHLLLRGFYL 237  
 Db 233 PNVLGESSVIDEHEPWLTLMLRL 256

## RESULT 8

QY2F54 PRELIMINARY; PRT; 259 AA.  
 ID QY2F54  
 AC QY2F54;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Hypothetical protein lin0253.  
 GN LIN0253.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tisseret A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of Listeria species.";  
 Science 294:849-852(2001).  
 RL EMBL; AL596164; CAC93486.1; --  
 DR ListList; LIN00253; --  
 DR InterPro; IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs; TIGR00671; baf; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 14.3%; Score 181; DB 16; Length 259;  
 Best Local Similarity 27.0%; Pred. No. 6.5e-07;  
 Matches 72; Conservative 42; Mismatches 109; Indels 44; Gaps 12;  
 QY 1 MIFIAVGNTRTLIAHPTHG-----VHFDSVSVATSLPPTTEILOQPGLTWLSAP 48  
 Db 2 ILVIDGNTCTGVYVYKQKLLKHWMTTDR-HRTSDELGMV-----LNFFSYA 50  
 QY 49 NREP-----VALGGVVPVPAALAAWREALATAEVR-----EP---DPGFFRAVPHDYHPPES 96



```
Db 51 NLTSDIOGIITISSVVPIM-----HAMETMCVRVENIRPLIVPG-IKTGLNLKVDNPRE 105
Qy 97 LGFRRCLLAAAMDYPCQDSIVIDMGTAITDILL-AGHFRRGRILPGIAMSRLRHEG 155
Db 106 IGSRIIVNAVAASEY-GTPVIVDFGTATFCYIDEAGVYQGAIAPIGIMISTEALYNR 164
Qy 156 TALLPEVYNAPAEIMGNDTSNAIOAGVIHLFADALRGCAITDFRQYSPQARILLITGGDAE 215
Db 165 AAKLPVDIAESSOLIGKSTVSMQAGIFYGICQEGIIAEMKQSNTPSPVVATGGGLA 224
Qy 216 RW--OPGIAGSLYQPHLLRGLFWIR 240
Db 225 RMITEKSSAVDILDPFLTKGLELLYR 251

RESULT 9
ID O32514 PRELIMINARY; PRT; 212 AA.
O32514;
01-JAN-1998 (TREMBlrel. 05, Created)
01-JAN-1998 (TREMBlrel. 05, Last sequence update)
01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Smiliar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MIYAZAKI;
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;
RT "Sequence analyses of two ferredoxin genes and their flanking regions
RT from Desulfovibrio vulgaris (Miyazaki F).";
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
DR EMBL; AB005550; BAA21476.1;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR QY SEQUENCE 212 AA; 22336 MW; 23E1789973A34D7 CRC64;

Query Match 14.3%; Score 180.5; DB 2; Length 212;
Best Local Similarity 29.4%; Pred. No. 5.6e-07;
Matches 63; Conservative 31; Mismatches 83; Indels 37; Gaps 9;

Qy 6 VGNTRTLAHTDGVHDSVATSLPPTTEILOQPLTWSAPNR--EPVALGGVWPAAL 63
b 10 IGNTNVKI-----GIAVETAVLTSYVLPTD---PGQTDDSIGLRLLEVLRHAGLGPADV 60
Qy 64 AAWREALATAEVREPDGFFRR-----AVPHD--YHPPESLGFDRCC 104
Db 61 GA---CVASSVWGVNP-LRRACERYLYRKLFPAGDIAIPLDNRYERPAAEVGADRLVA 116
Qy 105 LLAAMDYPGODSIV-IDMGTAITDILLAGHFRRGRILPGIAMSRLRHEGTALLPEV 163
Db 117 AYAAARLYPGRSLVSYDFGTATFCYIDEAGVYQGAIAPIGIMISTEALYNR 176
Qy 164 LNAPAE--MLGNDTSNAIOAGVIHLFADALRGAI 195
Db 177 LEVEEDSPVIGRSTTSLNHGFIQFAAMTEGVL 210

RESULT 10
ID Q9F985 PRELIMINARY; PRT; 258 AA.
AC Q9F985;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V;
RA Vasquez C., Pichuanes S., Saavedra C.;
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198621; AAG28531.1;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR QY SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 14.3%; Score 180.5; DB 2; Length 258;
Best Local Similarity 27.9%; Pred. No. 7.1e-07;
Matches 72; Conservative 43; Mismatches 106; Indels 37; Gaps 14;

Qy 1 MIFIA-VGNTRTLAHTDGVHDSVVS-----VATSLPPTTE-----ILOQPLTWS 46
Db 1 MIFVLDVGNVTIVLG-VYDG---DELKHHWRLETSGKTEDEYGMTIKALLNHVGLQF-- 54
Qy 47 APNREPVALGGVWPAAL-AAWREALATAEVRE--PDGFFRRAVPHDYHPPESLGFDRCC 103
Db 55 -SDIDGIISSVVPPIFMFALERMCKLYFIKPIIVGPG-IKTGLNKKYDNPREVGADRI 112
Qy 104 CLLAAAMDYPGODSIVIDMGTAITDILL-AGHFRRGRILPGIAMSRLRHEGTALLPEV 162
Db 113 NAVAGIHLV-GSLIIVDFGTATTCYINEHKQYMGGAIPAGIMISTEALFARAALPRI 171
Qy 163 VLNAPAEMLGNDTSNAIOAGVIHLFADALRGAITDFRQYSP-QARILLTGGDAERWQPGI 221
Db 172 EIAREDDLIIGKNTVSAMQAGILYGVQGVGIVSRMKAKSPVPKVIATGGLASL----I 227
Qy 222 AG-----SLYOPHLLLRG 234
Db 228 ASESNDIVIDVDPFLTLTG 245

RESULT 11
QY2M4 PRELIMINARY; PRT; 295 AA.
AC Q8Y2M4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative transcription regulation accessory factor transcription
DE regulator protein.
GN RSG0311 OR RS03278.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GMI1000;
RX MEDLINE-21681879; PubMed-11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Bottier P., Camus J.C., Cattolico L.,
RA Chandlee M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646058; CAD13839.1;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR Complete proteome.
DR QY SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 14.2%; Score 180; DB 16; Length 295;
Best Local Similarity 26.2%; Pred. No. 9.2e-07;
Matches 73; Conservative 30; Mismatches 98; Indels 78; Gaps 10;

Qy 1 MFIATVGNTRTLAHTDGVHDSVATSLPPTTEILOQPLTWSAPNRREPVALGGVVP 60
```

```
Db 13 LLLIDAGNTRIKWNT-----ADVAPPA--VAPGGTPWQHAGARPHDQLAELY- 59
QY 61 AALAAWREALATAEYREPDGCFRRVAP-----
Db 60 ---EDWRCHAGAGMAPDVMWISVVGAPALDCAICARVFDGARLRIVASEAAAAAGLR 116
QY 89 HDYHPESLGRDRCCLLAAMNDYGDQSVIVDMGTATIDLLA-GGHRGGRILPGIAM 147
Db 117 NGYRDPAQIGTDWVGAVGARHAWPDATLLLVTTAGTATATLDIVAPDGRFAGGLILPGLTL 176
QY 148 SLRGLHEGTALLPEVVL-----NAPAEH--LCNDTSNAIQAGVIHLFADALRGAT 196
Db 177 MMRALSRTAQLPEIDIGYLAARDQAAPADVPWSHADNTQDAIALGCV-----TAQAGAIA 232
QY 197 DF-----ROYSPQARILITGGDAERWQPGIAGSLYOPHL 230
Db 233 QTWQALQAQYPCPYRCVLSGG-----ARALAPHL 262

RESULT 12
Q9KGH5 PRELIMINARY; PRT; 254 AA.
AC 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
EX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001507; BAB03805.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 14.2%; Score 179.5; DB 16; Length 254;
Best Local Similarity 26.1%; Pred. No. 8.4e-07;
Matches 61; Conservative 47; Mismatches 97; Indels 29; Gaps 9;

QY 1 MIFTAVGNTRILLATHDGVHDSV-----ATSLPTE-----ILQOQGLTWLS 46
Db 2 ILVIDGNTNTVL-----GVYQDETIVHWHRLATSRKQTEDEYATVRSFLFDHAGLQF-- 54
QY 47 APNREPVALGGVWPAALAAWREALATAEYREP---DPGFFRRAVPHDYHPESLGFDRRC 103
Db 55 -ODIDGIVSSVPPMPFSLQMKKYPHTVPMITGPG-ITGLNIRKYNPNKEVGAD-RI 111
QY 104 CLLAAMNDYGDQSVIVDMGTATIDLL-AGHFRGGRILPGIAMSRLRGEGTALLPEV 162
Db 112 VNVAATIELYGPALVDFGTATYCLINEKQYAGGVAPGIMISTEALYHRASKLPRI 171
QY 163 VLNPAEMLGNDTSNAIQAGVIHLFADALRGATDFR-QYSPQARILITGGDAE 215
Db 172 EIAKPKQVGTNTIDSMQSGIFYGVVSQDGVVKKRMAQAESEPKVIATGGLAK 225

RESULT 13
P74045
```

```
ID P74045 PRELIMINARY; PRT; 257 AA.
AC P74045;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein slr0812.
GN slr0812.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=13 / TYPE A;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugliara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18120.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 257 AA; 27702 MW; 4291E51EE91FB4A2 CRC64;

Query Match 13.8%; Score 174; DB 16; Length 257;
Best Local Similarity 27.9%; Pred. No. 2.4e-06;
Matches 60; Conservative 34; Mismatches 73; Indels 48; Gaps 8;

QY 40 PGLTWLSAPN-----REPVALGGVWPAALAAWREALATAEYREPDPGFFRRAVP 88
Db 40 PLQTTWTDYNPKSAQLPVLVLLGKVPMLASVVPQTEVWR-----VYQPKILTLKNLPL 92
QY 89 HDYHPESLGRDRCCLLAAMNDYGDQSVIVDMGTATIDLLAGHFR-----GGRIPL 143
Db 93 VNLYP--SFGIDRALAGLGTLY-GFPCIVVDGTAITIT-----GFDQDKLVGGAILP 145
QY 144 GTAMSLRGHEGTALLPEVVLNAPAEMLGN---DTSNAIQAGVIHLFADALRGATDPRO 200
Db 146 GIGLQATLGDRLLALPKLEMDQLTELPDRWALDTPSAIFSGVGVYGVGLAQSLYQDWQR 205
QY 201 YSPQARILITGGDAERWQPGIAGSLYQPHLLLRGF 235
Db 206 LFPGAAMVITGGDK-----ILHGF 225

RESULT 14
Q8XHL5 PRELIMINARY; PRT; 259 AA.
ID Q8XHL5;
AC Q8XHL5;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=13 / TYPE A;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB82174.1; -
```

-DR InterPro: IPR004619; Baf.  
DR InterPro: IPR000515; BPD\_transp.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs: TIGR00671; baf; 1.  
DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
-KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match 13.7%; Score 173.5; DB 16; Length 259;  
Best Local Similarity 25.9%; Pred. No. 2.6e-06;  
Matches 60; Conservative 42; Mismatches 109; Indels 21; Gaps 7;  
QY 1 MIFTAVGNTRTLLAHTHGVHDSVSVATSLPPT-----EILQPGTLTWSAPNRPVA 54  
DB 2 ILLIDVGNINVLGHNDKXIASWRISTSKTSDEYSIQVMQFNQAKLPEDVEGII 61  
QY 55 LGVVVPAALAAWREALATAEVREP---DPGFRAVPHDYHPPEISLGFDRRCCLAAAMD 111  
DB 62 ISSVVPNTMHSLENNVRKCFCKEPIVVGPG-IKTGINIKYDNPKEVGAD-RIVNAVAAFE 119  
QY 112 YPGQDSIVIDMGTAITI-DLLAGHFRGRILPGTAMSLRGLHEGTALLPEVVLNAPAE 170  
DB 120 KHKPMIITDGTATFTCAITEKGDYLGNCIPGQISADALFERAAKLPRIELEKPKSV 179  
QY 171 LGNDSNAIQGVH-----LFADALRGAITDFRQYSPQARILITGGDAE 215  
DB 180 ICKNTVTSMQAGIYIGKVEYIVKRMKEMMDLGEKEP--FVLATGGLAK 229

RESULT 15  
Q9CD56 PRELIMINARY; PRT; 274 AA.  
AC Q9CD56;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein ML0232.  
GN ML0232.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RA MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
DR EMBL; AL583917; CAC29740.1; -.  
DR Leproma; ML0232; -.  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs: TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 13.6%; Score 172; DB 16; Length 274;  
Best Local Similarity 32.0%; Pred. No. 3.7e-06;  
Matches 54; Conservative 26; Mismatches 77; Indels 12; Gaps 5;  
QY 79 DPGFFRAVPHDYHPPEISLGFDRRCCLAAAMDYPGQDSIVIDMGTAITDILL-AGGHER 137  
DB 91 EPG-VRTGIPLLVDPNKEVGADRIVNCIAAFHKF-GQAAIVVDVGSSICVDVSAKGEFL 148

QY 138 GGRILPGIAMSLRGLHEGTALLPEVVLNAPAEMLGNDTSNAIQAGVILHFDALRGAITD 197  
DB 149 GGAIAPGVQVSSDAAARSAAALRRVELARPSRVVGNKTVTECMQAGVVFAGLVGLVGR 208  
QY 198 FRQ-----YSPQARILITGGDAERWQPGI-AGSLYQPHLLLRGFYL 237  
DB 209 MEQDVEEFGSDLGNGRVAVVATGHTAPLLPELHTVDHYDRHLTLHGRL 257

Search completed: June 24, 2003, 22:16:22  
Job time : 22.1995 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 9.77804 Seconds  
(without alignments)  
725.188 Million cell updates/sec

Title: US-09-813-453A-63  
Sequence: 1 MIFTAGVTRILLATHDGV.....AGSLYQPHLLRGLFYNIWRG 241

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

earched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.5	7.8	6095	4	US-09-144-085-2
2	94.5	7.5	5215	4	US-09-105-537-2
3	92	7.3	5087	4	US-09-144-085-1
4	86.5	6.8	3724	2	US-08-804-227C-10
5	86.5	6.8	3724	2	US-08-804-198-4
6	86	6.8	1077	4	US-09-412-210-1
7	86	6.8	1864	2	US-08-804-227C-3
8	85.5	6.8	4545	2	US-08-804-227C-14
9	85.5	6.8	4550	2	US-08-804-227C-8
10	85.5	6.8	4550	2	US-08-804-198-2
11	82	6.5	654	4	US-09-620-412C-341
12	82	6.5	1752	4	US-09-556-877-180
13	82	6.5	1752	4	US-09-620-412C-180
14	80	6.3	3567	2	US-07-642-734C-4
15	80	6.3	3567	3	US-08-439-009A-4
16	78.5	6.2	1399	4	US-08-462-467B-14
17	78	6.2	1958	1	US-07-945-283-2
18	77.5	6.1	3816	4	US-09-428-517-3
19	77	6.1	485	4	US-09-384-212-2
20	77	6.1	1317	3	US-09-083-521-7
21	77	6.1	1891	2	US-08-804-227C-12
22	77	6.1	1891	2	US-08-804-198-6
23	76	6.0	396	4	US-09-046-992-4
24	76	6.0	420	1	US-08-391-259-2
25	76	6.0	420	1	US-08-391-259-7
26	76	6.0	420	1	US-08-391-259-10
27	76	6.0	420	1	US-08-391-259-11

28 76 6.0 420 2 US-08-839-425-2 Sequence 2, Appli  
29 76 6.0 420 2 US-08-839-425-7 Sequence 7, Appli  
30 76 6.0 420 2 US-08-839-425-10 Sequence 10, Appli  
31 76 6.0 420 2 US-08-839-425-11 Sequence 11, Appli  
32 76 6.0 456 1 US-08-021-601-6 Sequence 6, Appli  
33 76 6.0 456 1 US-08-082-849B-6 Sequence 6, Appli  
34 76 6.0 456 5 PCT-US94-01624-6 Sequence 6, Appli  
35 76 6.0 472 1 US-08-021-601-8 Sequence 8, Appli  
36 76 6.0 472 1 US-08-082-849B-8 Sequence 8, Appli  
37 76 6.0 472 5 PCT-US94-01624-8 Sequence 8, Appli  
38 76 6.0 508 1 US-08-021-601-10 Sequence 10, Appli  
39 76 6.0 508 1 US-08-082-849B-10 Sequence 10, Appli  
40 76 6.0 508 5 PCT-US94-01624-10 Sequence 10, Appli  
41 76 6.0 599 1 US-08-463-163-3 Sequence 3, Appli  
42 76 6.0 613 1 US-08-405-615-1 Sequence 1, Appli  
43 76 6.0 613 2 US-08-461-234-1 Sequence 1, Appli  
44 76 6.0 613 2 US-08-463-480-1 Sequence 1, Appli  
45 76 6.0 613 4 US-09-479-479-2 Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-09-144-085-2  
; Sequence 2, Application US/09144085  
; Patent No. 6280999  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Ashley, Gary  
; APPLICANT: Julien, Bryan  
; APPLICANT: Ziermann, Rainer  
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
; FILE REFERENCE: 30062-20020.20  
; CURRENT APPLICATION NUMBER: US/09/144,085  
; CURRENT FILING DATE: 1998-08-31  
; EARLIER APPLICATION NUMBER: 09/010,809  
; EARLIER FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 6095  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-144-085-2

Query Match 7.8%; Score 98.5; DB 4; Length 6095;  
Best Local Similarity 26.0%; Pred. No. 0.73;  
Matches 58; Conservative 21; Mismatches 79; Indels 65; Gaps 10;  
QY 11 TLLAHFDGVDVSVSATSLPTEILQOPLTWLSAPNRPVALGVVPAALAAWREAL 70  
Db 1293 TALGOSQGV---ASLSHEATRQALSLLQ-----AWLSEPRLDVAVELVWVTRGAVGAAPDDA 1346  
QY 71 ATAEEVPEPPGPFERRAVPHDYHPPESELGDFDRRCCLLAAMDYPQDSIVIDMGTAITID 129  
Db 1347 VQDLARAPLWGLVR-----AARSEHPEERLRLIDVGTPEVDAG 1384  
QY 130 LAGGHFRGRRLPGIAMSRLGHEGTALLPEVVLNAPAEMLGNDTSTNAIOAGVHIHFAD 189  
Db 1385 LLA-----RAL--ATAEPELALRGGA-----ALAAALVRAQAA 1416  
QY 190 A---LRGATDFRQYSPQARILITGGDAERWQGIAGSLYQPH 229  
Db 1417 AEELTRGA----RELDPACTVLVTGTGTGELGQ-AIAAHLVRAH 1454

RESULT 2  
US-09-105-537-2  
; Sequence 2, Application US/09105537A  
; Patent No. 6265202

Db	4532	WLEPSEPLEATELVNITRGAAGVGAAPDDDAVEDALAPLWGLVR	-----	457
QY	104	CLLAAAMDPGQDSIVIDMGT-AITIDLLAGGHFRGRILPGIAMSRLGHEGTALLPEV	162	
Db	4573	---AARSEHPERGLRLMDVGTPEVDAGLLA	-----	4610
QY	163	VLNAPAEMLGNDTSNAIQAGVYHLFADALRGAITDFROYSPQARILITGQDAERWQPIA	222	
Db	4611	ALRGGG-----ALAAARLVR--AQAVAEELTRARELDPAGTGLVLTGTGELGQ-AVA	4658	
QY	223	GSLYQPH 229		
Db	4659	AHLVRAH 4665		
<p>RESULT 4</p> <p>US-08-804-227C-10</p> <p>Query Match 7.5%; Score 94.5; DB 4; Length 5215;</p> <p>Best Local Similarity 27.8%; Pred. No. 1.6;</p> <p>Matches 72; Conservative 23; Mismatches 99; Indels 65; Gaps 18;</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Sherman, D.H.</p> <p>APPLICANT: Liu, H.</p> <p>APPLICANT: Xue, Y.</p> <p>APPLICANT: Zhao, L.</p> <p>TITLE OF INVENTION: DNA encoding methymycin and pikromycin</p> <p>FILE REFERENCE: 600.4380S1</p> <p>CURRENT APPLICATION NUMBER: US/09/105,537A</p> <p>CURRENT FILING DATE: 1998-06-26</p> <p>NUMBER OF SEQ ID NOS: 43</p> <p>SOFTWARE: FastSeq for Windows Version 3.0</p> <p>SEQ ID NO 2</p> <p>LENGTH: 5215</p> <p>TYPE: PRT</p> <p>ORGANISM: Streptomyces venezuelae</p> <p>US-09-105-537-2</p>				
Y	10	RTLLAHTHDGVHFDTSVATSLPPTTILQ-----QPGTWSAPNREPVAL-----G	56	
Db	1833	RTLAA-----QDATVVEIGPDVAVLTALAEALAPG---TDAPDARDVTVPVLLRAG	1881	
QY	57	GVVPAALAAWREALATAEYRE-----PD-----PGF-FRRVAVPHDYHPE-----	95	
Db	1882	REPETLAA---GLATAHVHGAPLDRASFPGDGRDLDTYAFRR---EHWLTPERTDA	1936	
QY	96	SIGFTD-RRCCLIAAMDPGQDSIVIDMGTAITIDLLAGGHF-RGRILPGTAM-----S	148	
Db	1937	RALGFDPAHPRLITTTVEVAGDGVLLTGLSLTQDPLWLDHVMNGAVLLPATAFLELAL	1996	
QY	149	LRGLHEGTALLPEVLNAPAEMLGNDTSNAIQAGVYHLFADALRGAITDFROYSPQARI	207	
Db	1997	AAGDHVGAVRVEELTLEAPL-VLPERGAVRIQGV---SGDGESPAGRTFGVYSTPDSG-	2051	
QY	208	LITGDAER-WOPGIAGSL 225		
Db	2052	DTGDDAPREWRHVSGVL 2069		
<p>RESULT 3</p> <p>US-09-144-085-1</p> <p>Sequence 1, Application US/09144085</p> <p>Patent No. 6280999</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Gustafsson, Claes</p> <p>APPLICANT: Betlach, Mary C.</p> <p>APPLICANT: Ashley, Gary</p> <p>APPLICANT: Julien, Bryan</p> <p>APPLICANT: Ziermann, Rainer</p> <p>TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA</p> <p>TITLE OF INVENTION: THEREFOR</p> <p>FILE REFERENCE: 30062-20020.20</p> <p>CURRENT APPLICATION NUMBER: US/09/144,085</p> <p>CURRENT FILING DATE: 1998-08-31</p> <p>EARLIER APPLICATION NUMBER: 09/010,809</p> <p>EARLIER FILING DATE: 1998-01-22</p> <p>NUMBER OF SEQ ID NOS: 8</p> <p>SOFTWARE: PatentIn Ver. 2.0</p> <p>SEQ ID NO 1</p> <p>LENGTH: 5087</p> <p>TYPE: PRT</p> <p>ORGANISM: Sorangium cellulosum</p> <p>US-09-144-085-1</p>				
Y	44	WLSAPNREPVALGGVVPAAALAAWREALATAEYREPDGFFRRAVPHDYHPPESLGRDRC	103	
Db	45	Conservative 21; Mismatches 67; Indels 54; Gaps 7;		
QY	44	WLSAPNREPVALGGVVPAAALAAWREALATAEYREPDGFFRRAVPHDYHPPESLGRDRC	103	
<p>Query Match 7.3%; Score 92; DB 4; Length 5087;</p> <p>Best Local Similarity 24.1%; Pred. No. 2.9;</p> <p>Matches 45; Conservative 21; Mismatches 67; Indels 54; Gaps 7;</p>				

Db	4532	WLEPSEPLEATELVNITRGAAGVAPDDDAVEDALAPLWGLVR	-----	457
QY	104	CLLAAAMDPGQDSIVIDMGT-AITIDLLAGGHFRGRILPGIAMSRLGHEGTALLPEV	162	
Db	4573	---AARSEHPERGLRLMDVGTPEVDAGLLA	-----	4610
QY	163	VLNAPAEMLGNDTSNAIQAGVYHLFADALRGATIDFROYSQPARILITGDAERWQPGIA	222	
Db	4611	ALRGGG-----ALAAALVR--AQAVAEELTRARELDPACTVLVTGTTGELGQ-AVA	4658	
QY	223	GSLYQPH 229		
Db	4659	AHLVRAH 4665		
<p>RESULT 4</p> <p>US-08-804-227C-10</p> <p>Query Match 6.8%; Score 86.5; DB 2; Length 3724;</p> <p>Best Local Similarity 25.8%; Pred. No. 7.5;</p> <p>Mismatches 71; Conservative 21; Mismatches 102; Indels 81; Gaps 14;</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Dehoff, Bradley S.</p> <p>APPLICANT: Kuhstoss, Stuart A.</p> <p>APPLICANT: Rostock, Paul R., Jr.</p> <p>APPLICANT: Sutton, Kimberly L.</p> <p>TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES</p> <p>NUMBER OF SEQUENCES: 15</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: THOMAS G. PLANT 1501</p> <p>STREET: LILLY CORPORATE CENTER</p> <p>CITY: INDIANAPOLIS</p> <p>STATE: IN</p> <p>COUNTRY: USA</p> <p>ZIP: 46285</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM Compatible</p> <p>OPERATING SYSTEM: MS-DOS</p> <p>SOFTWARE: ASCII(DOS) Text only</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/804, 227C</p> <p>FILING DATE: February 21, 1997</p> <p>CLASSIFICATION: 435</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Plant, Thomas, G.</p> <p>REGISTRATION NUMBER: 35,784</p> <p>REFERENCE/DOCKET NUMBER: X-8231</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 317-276-2459</p> <p>INFORMATION FOR SEQ ID NO: 10:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 3724 amino acids</p> <p>TYPE: amino acid</p> <p>TOPOLOGY: unknown</p> <p>MOLECULE TYPE: peptide</p> <p>US-08-804-227C-10</p>				
QY	12	LLAHTHDGVHFDVSVSATSLPPTTEILQQPGLTWLSA--PNREP----	ALGGVVPAALAAW 66	
Db	3044	LLAPVPDVGW---SYAQAAASVPAVFLSAYGLVTLGLRPGERVVLHAAAGVGVMAAVOIA	3100	
QY	67	R-----EALATAEVRPDPG---FFRAVPHDHYPPESLGFDRCCLLAAAMDPGQDSIV	119	
Db	3101	RHLGAEVLTATA-----SPGKWDALRAMGITDDHLSASRTIDFATAFTGA--DGTSRADV	3153	
QY	120	IDMGTAITDILLAGGHFRGRIL	-----	PGIAMSRLGHEGTALLPEV 162
Db	3154	LSLTKFEFVDSIGLLRPGGRLELKGKTVDRPDERTAAEHPGVRYRAFDLINE	-----	ADALRGATIDFREQ 200
QY	163	VLNAPAEMLG-----NDTSNAIQAGVYHLF	-----	457

```
Db 3206 ---AGPDALGRLRLMDLFAAGVHLPLPVVTHDVRRAADALR-TISOARHTGKLVLTMP 3261
Qy 201 --YSPQARILITGDAERWOPGTAGSLYQPHLLR 233
Db 3262 PAWHPYGTVLVTGG-----TGALGSRIRARHLASR 3290

RESULT 5
US-08-804-198-4
; Sequence 4, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3724 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-198-4

Query Match 6.8%; Score 86.5; DB 2; Length 3724;
Best Local Similarity 25.8%; Pred. No. 7.5;
Matches 71; Conservative 21; Mismatches 102; Indels 81; Gaps 14;

Qy 12 LLAATHGVHDPSVATSLPPTTEILOQPGTWSLA--PNRPV---ALGGVVPAAALAAW 66
Db 3044 LLAPVPDGV---SYAQAASVPAVELSAYIGLVTLAGLPGERVLVHRAAGVGMAAVQIA 3100
Qy 67 R----EALATAEVEPDGP---FFRRAVPHYHPPESLGFDRCCLLAAAMDYPGQDSIV 119
Db 3101 RHLGAELATA-----SPGKWDALRAMGITDDHLASSRTLDFAFTGA--DGTSRADV 3153
Qy 120 IDMGTAITDILLAGHFRGGRII-----PGIAMSRLGLHEGTALLPEV 162
Db 3154 LNSITKEFVDASLGLRPGGRFLGLGKTDVDRDPERIAAEHGVRYRAFDLNE----- 3205
Qy 163 VLNPAEMLG---NDTSNAIQAGVTHLF-----ADALRGAITDFRQ----- 200
Db 3206 ---AGPDALGRLRLMDLFAAGVHLPLPVVTHDVRRAADALR-TISOARHTGKLVLTMP 3261
Qy 201 --YSPQARILITGDAERWOPGTAGSLYQPHLLR 233
Db 3262 PAWHPYGTVLVTGG-----TGALGSRIRARHLASR 3290
```

```
RESULT 6
US-09-412-210-1
; Sequence 1, Application US/09412210
; Patent No. 6403358
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 21529, A NOVEL ADENYLATE CYCLASE
; FILE REFERENCE: 5800-47
; CURRENT APPLICATION NUMBER: US/09/412,210
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1077
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-412-210-1

Query Match 6.8%; Score 86; DB 4; Length 1077;
Best Local Similarity 25.9%; Pred. No. 1.4;
Matches 58; Conservative 23; Mismatches 66; Indels 106; Gaps 16;

Qy 32 PPTTEILOQPGTWSLAPNREP--VALGGVVPAAALAAWREALATAEYRE--PDGFFERRAV 87
Db 10 PPSEDLFYE--TYTSLSQYPLLLLLGLVLCALALL-AVAWASGRELTSDPSFLTTL 66
Qy 88 PHDYHPPESLGFDRCCLLAAAMDYPGQDSIVIDMGT-----ATTIDLLAG 133
Db 67 -----CAL-----GGFSLLLGLASRQRQRQRWTRPLSLGLVWVALLAL 103
Qy 134 GH---FRGG-----RILP-----GIAMSLR-----GLHEG-- 155
Db 104 GHAFLTGGVVSAMDQVSYFLVFTAYAMLPLGMRDAAVAGLASSLSHLLVLGLYLPQ 163
Qy 156 ----TALLPEVVLNAPAEMLGNDTSNAIQAGVTH--LFADALRGAITDFRQYSPQARILI 209
Db 164 PDSRPALLPQLAANAVLFCGN-----VAGVTHKALMERALRATFREALSLSHRRRL- 216
Qy 210 TGGDAERWQPGIAGSLYQPHLL 232
Db 217 ---DTEK-----KHQEHLLL 228

RESULT 7
US-08-804-227C-3
; Sequence 3, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1864 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-3

Query Match 6.8%; Score 86; DB 2; Length 1864;

Best Local Similarity 27.8%; Pred. No. 3.1;

Matches 45; Conservative 14; Mismatches 43; Indels 60; Gaps 9;

QY 24 SVSVATSLPPTTEILOOGLTWLSAPNREPVALGGVVVPAALAAWREALATAEVRDPDGGFF 83

Db 920 SLAAVEL-PTIAFQTHY-WDAP-----AAPAALPAGLD-----DAG-- 956

Y 84 RRAVPHDYHPPESLGFRRRCCLAAADYP-GQDSIVIDMGTAITIDLLAGGHFRGGRIL 142

Db 957 -----HP-----LLSAALDPGGRGVVTGALSAAALPWAADHSVHGRTVL 997

QY 143 PGIAAMSLRGLH-----EGTALLPE-----VVINAP 167

Db 998 PGTALLDLAALHAPRVGELTTEAPLVLPEDGVEVRLRVVLAEP 1039

#### RESULT 8

US-08-804-227C-14

; Sequence 14, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

; APPLICANT: DeHoff, Bradley S.

; APPLICANT: Kustoss, Stuart A.

; APPLICANT: Rostock, Paul R., Jr.

; APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THOMAS G. PLANT 1501

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII(DOS) Text only

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,227C

; FILING DATE: February 21, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Plant, Thomas, G.

; REGISTRATION NUMBER: 35,784

; REFERENCE/DOCKET NUMBER: X-8231

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4545 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-804-227C-14

Query Match

Best Local Similarity 6.8%; Score 85.5; DB 2; Length 4545;

Matches 63; Conservative 21.9%; Pred. No. 13;

Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;

QY 11 TLLAHTHDCGVHF-DSVSVATSLPPTTEILOOGLTWLSAPNREPVALGGVV---PAALAA- 65

Db 2604 SLASHLHDELFGPDSEAEPAAPVPM-----ADREPIAIVGMACRYPGGVASP 2654

QY 66 ---WREALATAEVRDPD-----GFERRAVPHDYHPPES 96

Db 2655 DDLWDLVAGDGHTLSFPADRGWDVEGLYDPEVPKGSYVREGGFLRSAEFD---AEF 2711

QY 97 LGFDRRCCLAAADYPGODSIVIDMGTAITIDLLAGGHFRGGR--ILPGI----- 145

Db 2712 FGISPR---EATAMD-PQORLLLETSWREALERAGIYVDSLRGTRTGVFSISQODYATQL 2767

QY 146 --AMSLRGLHGTALLPEVVLNAPAEMLG-----NDTSNATQAGVIHLFADALR---- 192

Db 2768 GDAADTYGGHVLGTIGSVISGRVAYALGLEGPALVDTACSSSLVALHLAVQSLRRGEC 2827

QY 193 -----GAI-----TDFROYSPQ-----ARILITGGDAERWQPGI 221

Db 2828 DLALAGGVTVMATPTVVFVEFSRQRLAADGRCKAFAGADGTAWAEGV 2875

#### RESULT 9

US-08-804-227C-8

; Sequence 8, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

; APPLICANT: DeHoff, Bradley S.

; APPLICANT: Kustoss, Stuart A.

; APPLICANT: Rostock, Paul R., Jr.

; APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THOMAS G. PLANT 1501

; STREET: LILLY CORPORATE CENTER

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII(DOS) Text only

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,227C

; FILING DATE: February 21, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Plant, Thomas, G.

; REGISTRATION NUMBER: 35,784

; REFERENCE/DOCKET NUMBER: X-8231

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4550 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-804-227C-8

Query Match

Best Local Similarity 6.8%; Score 85.5; DB 2; Length 4550;

Matches 63; Conservative 21.9%; Pred. No. 13;

Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;

QY 11 TLLAHTHDCGVHF-DSVSVATSLPPTTEILOOGLTWLSAPNREPVALGGVV---PAALAA- 65

Db 2609 SLASHLHDELFGPDSEAEPAAPVPM-----ADREPIAIVGMACRYPGGVASP 2659

QY 66 ---WREALATAEVRDPD-----GFERRAVPHDYHPPES 96



```
Db 2660 DDLWDLVAGDHTLSPPADRGWDVEGLYDPGPVPGKSYVREGGFLRSAAEFD---AEF 2716
Qy 97 LGFDRRCCLLAAMNDYPCQDSIVIDMTGTAITIDLLAGGHFRGR--ILPGI-----145
Db 2717 FGISPR---EATAMD-PQORLLLETSWALERAGIVPDSLGRTRGVFGISQQDYATQL 2772
Qy 146 --AMSLRGLHGTALLPEVVLNAPAEMLG-----NDTSNAIQAGVIHLFADALR-----192
Db 2773 GDAADTYGGHVLTTGLSGVISGRVAYALGLEGPALTVDVTACSSSLVALHLAVQSLRRGEC 2832
Qy 193 -----GAI-----TDFRQYSPQ-----ARILITGGDAERWQPGI 221
Db 2833 DLALAGGYVMATPTVFVFEFSRQGLAADRCKAFAGADGTAWAEGV 2880

RESULT 10
US-08-804-198-2
Sequence 2, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rosteck, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804.198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P5113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4550 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-198-2

Query Match 6.8%; Score 85.5; DB 2; Length 4550;
Best Local Similarity 21.9%; Pred. No. 13;
Matches 63; Conservative 27; Mismatches 105; Indels 93; Gaps 14;

Qy 11 TLIAHTDGVHF-DSVSVATSLUPTEILQPGTLWLSAPNRPVALGGV---PAALAA- 65
Db 2609 SLASHLHDLGPDSEAEPAAPTPVM-----ADEREPIAVGMACHYPCGVASP 2659
Qy 66 ---WREALATAEVRPDP-----GFFRRVPHDYHPES 96
Db 2660 DDLWDLVAGDHTLSPPADRGWDVEGLYDPGPVPGKSYVREGGFLRSAAEFD---AEF 2716
Qy 97 LGFDRRCCLLAAMNDYPCQDSIVIDMTGTAITIDLLAGGHFRGR--ILPGI-----145
Db 2717 FGISPR---EATAMD-PQORLLLETSWALERAGIVPDSLGRTRGVFGISQQDYATQL 2772
```

```
Qy 146 --AMSLRGLHGTALLPEVVLNAPAEMLG-----NDTSNAIQAGVIHLFADALR-----192
Db 2773 GDAADTYGGHVLTTGLSGVISGRVAYALGLEGPALTVDVTACSSSLVALHLAVQSLRRGEC 2832
Qy 193 -----GAI-----TDFRQYSPQ-----ARILITGGDAERWQPGI 221
Db 2833 DLALAGGYVMATPTVFVFEFSRQGLAADRCKAFAGADGTAWAEGV 2880

RESULT 11
US-09-620-412C-341
Sequence 341, Application US/09620412C
Patent No. 6448234
GENERAL INFORMATION:
APPLICANT: Steven P. Fling
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
DIAGNOSIS OF CHLAMYDIAL INFECTION
FILE REFERENCE: 210121.469C7
CURRENT APPLICATION NUMBER: US/09/620.412C
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 341
LENGTH: 654
TYPE: PMT
ORGANISM: Chlamydia trachomatis
US-09-620-412C-341

Query Match 6.5%; Score 82; DB 4; Length 654;
Best Local Similarity 21.2%; Pred. No. 19;
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;

Qy 4 IAVGNTRTLAHTHDGVHFDPSVATSLPPTTEILOQPGTLWLSAPNRPVALGGV-----58
Db 176 VLSVTENLVGKEGGHLAKTVNISN-----LKSQFSF--SNNKANSSTGTATTAS 225
Qy 59 VPAALAAWREALATAEVRPDPGFFRRVPHDYHPESLGFDRRC---CLLA---AAMDY 112
Db 226 APAAAAAALQAAAAAAPPSPATPTYSVGVVGGAIY-GEKVTFS-QCSGTCQFSNQAIIDNN 283
Qy 113 PGQDSIVIDMG-----TAITIDLLAGGH---FRGGRILPGIAMSRLGHGTALLPEVVL 164
Db 284 PSQSSINVQGGAIYAKTSLSIGSSDAGTSYIFSGNSVSGKSTGTGQIAGAIYSPVTTL 343
Qy 165 NAPAEMLGNDTS-----NAIQ-----AGY-----183
Db 344 NCPATFSNNTASIAATPKTSSDSSGNSIKDTIGGAIAGTAITLSGVSRFSGNTADLGA 403
Qy 184 IHLFADALRGAITDFRQYSPQARILITGGD--AERWQPGTAGSLYQPHLLRG 234
Db 404 IGTLANANTPSATSGSQNSITEKITLENGSFIFERNQANKRGAIYSPSISIKG 456

RESULT 12
US-09-556-877-180
Sequence 180, Application US/09556877
Patent No. 6432916
GENERAL INFORMATION:
APPLICANT: Probst, Peter
APPLICANT: Bhatia, Ajay
APPLICANT: Skelky, Yasir
APPLICANT: Fling, Steve
APPLICANT: Maisonneuve, Jeff
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
DIAGNOSIS OF CHLAMYDIAL INFECTION
FILE REFERENCE: 210121.469C5
CURRENT APPLICATION NUMBER: US/09/556.877
CURRENT FILING DATE: 2000-04-19
NUMBER OF SEQ ID NOS: 305
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 180
LENGTH: 1752
```

; TYPE: PRT  
; ORGANISM: Chlamydia  
US-09-556-877-180

Query Match 6.5%; Score 82; DB 4; Length 1752;  
Best Local Similarity 21.2%; Pred. NO. 7.8;  
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;

	QY	4	I A V G N T R T L I A H D G V H F S V S V A T S P L P E T I L Q O B G L T W L S A P N R E P V A L G V -----	58
			::::   :   :   :   :   :   :   :   :   :	
	D b	760	V S L V S T E N L Y K G G G L H A K T V I N S ------L K S G F S F- -S N N K A N S S T G V A T T A S	809
			: :  :   :   :   :   :   :   :   :   :	
	QY	59	V P A A L A A R E A L A T A E V R E P D C F E R R A V P H D Y H P P E S L G F D R R C ---C L L A --A M D Y	112
			:   :   :   :   :   :   :   :   :   :	
	D b	810	A P R A A A S L Q A A A A A P S P A T Y S G V G G A I Y -G E K V T F S -Q C S T C Q F S G N Q A I D N N	867
			:	
	QY	113	P G O D S I V D M G ----T A I T I D L L A G G H----F R G R I L P O I A M S L R G L H G E T A L L P E V I L	164
			:   :   :   :   :   :   :   :   :   :	
	D b	868	P S O S S L U N V O G G A I Y A K T S L S I G S D A G T S Y F S G N S V S T G K T T G O I A G G A I Y S P V T L	927
			:	
	-QY	165	N A P A E M L G N D T S ----N A I O ----AGV--	183
			:   :   :   :   :   :   :   :   :   :	
	O	928	N C P A T E S N N T A S I A T P K T S S E D G S G N S I K D T I G G A I A G T A I L S G V S R E S G N T A D L G A A	987
			:	
	QY	184	I H L F A D A L R G A I T D F R Q Y S P Q A R I L T G G B -A E R W Q P G I A G S I Y O P H L L R G	234
			:   :   :   :   :   :   :   :   :   :	
	D b	988	I G L A N A N T P S A T S G S O N I T E R I T L E N G S F I F E R N Q A N K R G A I Y S P V S I K G	1040

```

RESULT 13
US-09-620-412C-180
; Sequence 180, Application US/09G20412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Filing
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIA
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 180
; LENGTH: 1752
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-620-412C-180

```

```

Query Match      6.5%; Score 82; DB 4; Length 1752;
-Best Local Similarity 21.2%; Pred. No. 7.8;
Matches 62; Conservative 43; Mismatches 114; Indels 74; Gaps 13;

QY      4 IAVNTRTLLAHTDHGVHDSVSVAISLPTTEILOQPGLTWLSAPNRPVALGGV----- 58
      :: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      760 VLSLVTEVLVGKGGGLHAKTVNISN-----LKSQSF--SNNKANSSSTGVATTAS 809

QY      59 VPAALAAREALATAEVRPDPGFFRRAYPHDYPPEISLGFDRRC---CLLA---AMDY 112
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      810 APAAAAASLOAAAAAAPSPPATPTYSGVGGAII-GEKVTF--QCSGTQCFQSGNQAI DNN 867

QY      113 PGODSVIDMG-----TATIDLLAGGH---PRGGRI LPCIAMSLRGLHEGTALLPEVVL 164
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      868 PQSSLNVOGGAIYAKTSLISIGSSDAGTSYIFSGNSVSTGKSOTTQIAGGAIYSPVTVL 927

QY      165 NAPAEMLGNDTS-----NAIQ-----AGV----- 183
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      928 NCPATPSNNTASITATPKTSSDDGSSGNSIKDTIGGAICTATILSGVSFSGNTADIGAA 987

QY      184 IHLFADALRGATDFRQYSPQARILITGDD--AERWQPGIAGSLYOPHILLRG 234
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      988 IGTLANANTPSATSGSONSITEKITLENGSEFIFERNOANKRAIYSPSVSIKG 1040

```

RESULT 14  
US-07-734C-4  
; Sequence 4, Application US/07642734C  
; Patent No. 5824513  
; GENERAL INFORMATION:  
; APPLICANT: Katz, L  
; APPLICANT: Donadio, S  
; APPLICANT: Mcalpine, J B  
; TITLE OF INVENTION: Recombinant DNA Method for Producing  
; Erythromycin Analogs  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Edward H. Gorman  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; STREET: Park Rd  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/642,734C  
; FILING DATE: 17-JAN-91  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Danckers, Andreas M  
; REGISTRATION NUMBER: 32652  
; REFERENCE/DOCKET NUMBER: 4952.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-9396  
; TELEFAX: 708-938-2623  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3567 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-642-734C-4

Query Match	6.3%;	Score 80;	DB 2;	Length 3567;
Best Local Similarity	23.9%;	Pred. No. 37;		
Matches	58;	Conservative 13;	Mismatches 60;	Indels 112; Gaps 11;
QY	54	ALGGVVPAALA	---	AWREALATAEYREPDGCFRRRAVPHDYHPES---LGFDRCC--- 104
Db	2966	AAGVGMAAVALARRAGA	EVLTAGPAK--HGTLRALGLDDDEHIASSRETGFARKFRERT 3023	
QY	105	-----	LLAAADYPGQDSIVDMGTAITDLLAGHFRG----- 138	
Db	3024	GGRVDVVLNSLTGCLLDESADLLAEDGVFVEMGKT---	DLRDAGDFRGYAPDFDLGEAG 3080	
QY	139	---GRI---	---LPGIAMS---	RGLEHGTAALLPEVLNAPA 168
Db	3081	DDRLGEILREVVCLLGAGELDRLPYSAWELGSAQAQHWSRGRHVG---	KLVLTPQA 3135	
QY	169	EMIGNDTSNAIQAGVTHLFADALRGAITDPROYSPQARILITGGDAERWQPGIAGSLYOP 228		
Db	3136	PV-----	-----DPDGTVLITGG-----	TGTLGRLLAR 3158
QY	229	HLL 231		
Db	3159	HLV 3161		

RESULT 15  
US-08-439-009A-4  
; Sequence 4, Application US/08439009A  
; Patent No. 6004787  
; GENERAL INFORMATION:

APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven F. Weinstock  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3567 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-439-009A-4

Query Match 6.3%; Score 80; DB 3; Length 3567;

Best Local Similarity 23.9%; Pred. NO. 37;

Matches 58; Conservative 13; Mismatches 60; Indels 112; Gaps 11;

QY 54 ALGGVVPAALA-----AWREALATAEVREPDGFFFRVPHDYHPPE--LGFDRCC--- 104

DB 2966 AAGGVGMAVALARRAGAEVLATGPAK--HGTLRALGLDDDEHIASSRETGFARKPRET 3023

QY 105 -----LLAAAMDYPGODSIVDMGTAITIDLLAGGHFRG----- 138

DB 3024 GGRGVYVINSITGELLDESADLLAEDGVFVEMGKT---DLRDAGDFRGYAPFDLGEAG 3080

QY 139 -----GRI-----LPGIAMS-----RGLHEGTALLPEVVLNAPA 168

DB 3081 DDLRGLLEVVVLLGAGELDRLPVSAWELGSAPALQHMRSRHHV-----KLVLTPA 3135

QY 169 EMLGNDTSNAIQAGVHILFADALRGATIDFRQYSQARILITGGDAERWQPGIAGSLYOP 228

DB 3136 PV-----DPDGTVLITGG-----TGTGLRLLAR 3158

QY 229 HLL 231

DB 3159 HLV 3161

Search completed: June 24, 2003, 22:10:51

Job time : 11.778 secs

**THIS PAGE BLANK (USPTO)**





\*ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1102

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <GLA>

A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0221

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match

Best Local Similarity 14.6%; Score 185; DB 2; Length 259;

Matches 72; Conservative 43; Mismatches 108; Indels 44; Gaps 12;

1 MIFAVGNT-----RTLLAH---THDGVHFDVSVSATSLPPTTEILOQPGTLTWSAP 48

2 ILVIDGNTCTGVYEKQKLLKHWRTTDR-HRTSDELGNTV-----LNFFSYA 50

49 NREP-----VALGVVPAALAAWREALATAEVR-----EP-----DPGFRRVAPVPHDYHPPE 96

51 NLTPSDIQGIISVVPIM-----HAMETMCVRVFNIRPLIVGPG-IKTGLNLKVDNPRE 105

97 LGFDRRCCLAAAMDYQDSIVDMGTATIDLL-AGGHRGGRGRIILPGIAMSRLGLHEG 155

106 IGSDRIVNAVAASEEY-GTPIVVVDFGTATFCYIDSGVYQGAIAPIGMISTEALYNR 164

156 TALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGATIDFROYSPQARILLTGDAE 215

165 AAKLPVDAIESQIIGKSVSSNOAGIFGVQCCEGITAEMKKQNASPVVATGGLA 224

216 RW--QPGIAGSLYQPHLLRGFYLIWR 240

225 RMITEKSSAVDILPFLTKGLELYR 251

RESULT 6

T36391

hypothetical protein SCE94.31c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000

C:Accession: T36391

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: Z21573

A:Accession: T36391

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-265 <OLI>

A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCE94.31c

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match

Best Local Similarity 14.6%; Score 184; DB 2; Length 265;

Matches 76; Conservative 39; Mismatches 113; Indels 36; Gaps 11;

1 MIFAVGNTTLLAHHDG-----VHFDVSVSATSLPPTTEILOQ-----QPGTLWSA 47

2 LLTIDVGNTHVLG-LFDGDEIVHWRISTDSRTADEL--AVLLQGLMGHMLGLGDELG 58

48 PNREPVALGVVPAALAAWREALATAEVRP-----DPGFRRVAPVPHDYHPPEGLFDRRC 103

59 DGIDGIAICATVPSVLHLEVRTRYGDPVAVLVEFG-VKTPGPIILTDHPKVGADRRII 117

104 CLLAAAMDYQDSIVDMGTATIDLL-AGGHRGGRGRIILPGIAMSRLGLHEGALLPEV 162

118 NAVAANELY--GGPAIVVDFGTATTFDAVSARGEYIGVIAPIGIEISVALGVKAQLRKI 176

Qy 163 VLNPAEMLGNDTSNAIQAGVIHLFADALRGAIT-----DFROYSPQARILLTGDAERWQ 218

Db 177 EVARPSVIGKNTVEAMQSGVIGFAGQVDGVVNRMARLADDDVDTVIATGGLA---- 232

Qy 219 PGIAGSL-----YQPHLLLRGFYL 237

Db 233 PNVLGESSVIDEHEFWLTMGLRL 256

RESULT 7

S66100

conserved hypothetical protein yacB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000

C:Accession: S66100; E69740

R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994

A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis c

A:Reference number: S65967; MUID:96051385; PMID:7584024

A:Accession: S66100

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <OGA>

A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g46

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; G

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma

Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scar

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S

keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchly

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshic

A:Authors: Yoshikawa, H.F.; zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: E69740

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <KUN>

A:Cross-references: GB:Z99104; GB:AL009126; NID:92632267; PIDN:CAB11846.1; PID:el182

A:Experimental source: strain 168

C:Genetics:

A:Gene: yacB

A:Start codon: TTG

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match

Best Local Similarity 14.3%; Score 181; DB 2; Length 233;

Matches 61; Conservative 34; Mismatches 102; Indels 22; Gaps 7;

1 MIFAVGNTTLLAHHDGVDHVSATSLPPTTE-----ILQPGTLTWSAPNRE 51

2 LLVIDVGNTHVLGVYHDGKLEYHWRIETSRKHTEDFCMILRSFLDHSGLMF---EQID 58

Qy 52 PVALGGVVPAALAAWREALATAEVRP---PGFFRRVAPVPHDYHPPEGLFDRRCCLAA 108

Db 59 GIITSSVPPINFALERCTRYFHIEPQIVGFG-MKTGLNIDKYNPKFKEYGADRIVNAVA 117

Qy 109 AMDYPGQDSIVDMGTATIT---IDLLAGGHFRGGRILPGIAMSRLGHEGTLTPVVLN 165

Db 118 IHLV-GNPLIVDFGTATYCYID--ENKYGGAIAPIGTTISTEALYSRAAKLPRIET 174

Qy 166 APAEMLGNDTSNAIQAGVIHLFADALRGAITDFROYSPQ 204

Db 175 RPDNLIGKNTVSAMQSGILFGYVGVEGIVKEMKWAQK 213





RESULT 15  
E97293  
probable t



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.84794 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453A-63

Perfect score: 1264

Sequence: 1 MIFTAVGNRTLLAHTHDGV.....AGSLYQPHLLLRGFLYINIRG 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181	14.3	233	1 YACB_BACSU	P37564 bacillus su
2	154.5	12.2	267	1 BAE_BORPE	Q45338 bordetella
3	96.5	7.6	557	1 PAC1_PSESV	Q05053 pseudomonas
4	94	7.4	469	1 LEU2_RHIME	Q92176 rhizobium m
5	92	7.3	532	1 CKX1_ORISA	Q91d66 oryza sativ
6	92	7.3	690	1 VG42_BPMU	Q9tlv6 bacterioph
7	91.5	7.2	583	1 XP65_MYCTU	Q50733 mycobacteri
8	91	7.2	488	1 DHAL_PSESP	P33008 pseudomonas
9	90	7.1	557	1 PAC1_PSES3	P15557 pseudomonas
10	89	7.0	469	1 LEU2_BRUME	Q8yjc9 bruceella me
11	88.5	7.0	980	1 PEX6_HUMAN	Q13608 homo sapien
12	86	6.8	400	1 ARGD_MYCTU	P94990 mycobacteri
13	86	6.8	561	1 BCHD_RHOCA	P26175 rhodobacter
14	86	6.8	732	1 TRPE_AZOBR	P50872 azospirillum
15	85	6.7	760	1 FTSH_MYCTU	P96942 mycobacteri
16	84.5	6.7	454	1 ATTY_HUMAN	P17735 homo sapien
17	84.5	6.7	1247	1 IRBP_HUMAN	P10745 homo sapien
18	84	6.6	286	1 AAC8_STRFR	P29809 streptomyc
19	84	6.6	1305	1 GNA_PRVKA	P97874 rattus norv
20	84	6.6	1733	1 VNQA_PRVKA	P33485 pseudorabie
21	83.5	6.6	525	1 P2BB_RAT	P20651 rattus norv
22	83	6.6	334	1 FEPD_ECOLI	P23876 escherichia
23	83	6.6	362	1 YD33_MYCLE	P53425 mycobacteri
24	83	6.6	629	1 YS50_MYCTU	Q05809 mycobacteri
25	82	6.5	1099	1 CARE_CAUCR	Q9a4d6 caulobacter
26	81	6.4	1754	1 PMPE_CHLTR	Q84418 chlamydia t
27	80	6.3	3567	1 LEU2_SACER	Q03132 saccharopol
28	79	6.2	469	1 SURE_RHIMO	Q98ef1 rhizobium l
29	78	6.2	247	1 SURE_THEMEA	P96112 thermotoga
30	78	6.2	691	1 DNLI_MYCTU	O53261 mycobacteri
31	77.5	6.1	424	1 THC2_METTH	O27617 methanobact
32	77.5	6.1	515	1 YJEF_ECOLI	P31806 escherichia
33	77.5	6.1	700	1 PURL_HALNI	Q9hr49 halobacteri

34	77	6.1	409	1 HUTL_RHILO	Q98311 rhizobium l
35	77	6.1	485	1 ANG1_HUMAN	P01019 homo sapien
36	77	6.1	637	1 MUTL_CAUCR	Q9rp66 caulobacter
37	77	6.1	1286	1 IRBP_BOVIN	P12661 bos taurus
38	76.5	6.1	1596	1 ACS2_ACEXY	Q59167 acetobacter
39	76	6.0	219	1 GPH_RHOCA	Q33512 rhodobacter
40	76	6.0	390	1 YL28_STRCO	P40181 streptomyc
41	76	6.0	391	1 ARGD_CORGL	Q59282 corynebacte
42	76	6.0	453	1 MTEL_RAT	O55171 rattus norv
43	75.5	6.0	449	1 TIG_RALSO	Q8xyp8 ralstonia s
44	75.5	6.0	463	1 YV30_MYCTU	O07035 mycobacteri
45	75.5	6.0	597	1 WD42_DICDI	P54686 dictyosteli

#### ALIGNMENTS

RESULT 1  
YACB\_BACSU STANDARD; PRT; 233 AA.  
ID YACB\_BACSU  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
[1]  
RN  
RP  
SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
[2]  
RN  
RP  
SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porvollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger T., Rivolta C., Roche E., Roche B., Rose M., Sadaie F.,  
RA Sato T., Scanlan E., Schleich R., Schroeter R., Scofield F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D26185; BAA05305.1; -
CC EMBL; Z99104; CAB11846.1; -
CC Subtilist; BG10133; yacB.
CC InterPro; IPR004619; Baf.
CC Pfam; PF03309; Bvg_acc_factor; 1.
CC TIGRfam; TIGR00671; baf; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;

Query Match      14.3%; Score 181; DB 1; Length 233;
Best Local Similarity 27.9%; Pred. No. 7.5e-09;
Matches 61; Conservative 34; Mismatches 102; Indels 22; Gaps 7;

1 MIFIAVGNRTLLAHTDGVHFDVSVSATSLPTE-----ILQPGLTWLSAPNRE 51
2 LLVIDVGNVTNVLGVYHDKLEHYHWRKTSRHKTEDEFGMILRSLDHSLMF---EQID 58
52 PVALGGVPAALAAWREALATAEVRPD---PCFFRRAYPHDYHPPESLGDFRRCCLAA 108
59 GIIISVVPPIFALERMCTKVFHTEPQIVGPG-MKTGLNPKYDNKEVGADRVNAVA 117
109 ANDYQDSIVIDMGTAIT---IDLLAGHFGGRILPGIAMSRLGLHGTALLPVLN 165
118 IHLV-GNPLVYDFVTFATTCYID--ENKQYMGGAIPGTTISTEALYSRAAKLPRIET 174
166 APAEMLGNDTSNAIOAGVIHLFPADALRGAITDFROYSPQ 204
175 RPDNIIGKTVSAMSGILFGVGVQEGIVKMKWQAKQ 213

RESULT 2
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
DE BAF.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OT Bordetella.
NCBI_TaxID=520;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshaizer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli.";
[1]
RN J. Bacteriol. 177:3801-3807(1995).
RN [2]
RN SEQUENCE OF 1-38 FROM N.A.
RC STRAIN-BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a bifa homolog in Bordetella pertussis.";
RT Submitted (JUL-1997) to the EMBL/Genbank/DDSIJ databases.
[3]
RN SEQUENCE OF 239-267 FROM N.A.
RC STRAIN-BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus

```

```

RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis.";
RL Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BYCAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; UI2020; AAA75361.1; -
CC EMBL; AF016461; AAC68834.1; -
CC EMBL; X90711; CAA62242.1; -
CC InterPro; IPR004619; Baf.
CC Pfam; PF03309; Bvg_acc_factor; 1.
CC Transcription regulation; Activator.
CC SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

Query Match      12.2%; Score 154.5; DB 1; Length 267;
Best Local Similarity 28.8%; Pred. No. 1.9e-06;
Matches 69; Conservative 27; Mismatches 109; Indels 35; Gaps 10;

1 MIFIAVGNRTLLAHTDGVHFDVSVSATSLPTEI----LQPG-LTWLSAPNREPVAL 55
2 ILLIDSGNSRLKVG-----WFDPAQAAREAPAFVDFNLDLALGRWLTLPFRPORA 55
56 GGVVPAALAAWREALATAEVRPDGPFRR-----RAVPHDYHPPESLGDFRRCCLAAAM 110
56 LGVNVAGLARGEAIATLRAGCGDIRWLRAQLAMGLRNGYRNPDLGADRACWACVGVLA 115
111 DYPG--QDSIVIDMGTAITIDLLAGH-FRGGRILPGIAMSRLGLHGTALLP---EVL 164
116 ROPSVHPPLLVASFSGTATTLDTIGPDNVFGLILPGFAMRGALAYGTALHPLADGLVA 175
165 NAPAEMLGNDTSNAIOAGVIHLFPADAL-RGAITDFROYSPQARILITGG-----DAER 216
176 DYP-I-----DTHQAIASGIAAQAAGAVRQWLAGRQRYGQAPETVYAGGGWPEVQEAER 230

RESULT 3
ID PAC1_PSESV STANDARD; PRT; 557 AA.
AC Q05053;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acylase ACY 1 [includes: Cephalosporin acylase (EC 3.5.1.-) (GL-7ACA
DE acylase); Gamma-glutamyltranspeptidase (EC 2.3.2.2) (GGT)].
DE ACII.
GN ACII.
OS Pseudomonas sp. (strain V22).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=33068;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-12 AND 367-380.
RX MEDLINE=93041922; PubMed=1358202;
RA Ishiye M., Niwa M.;
RT "Nucleotide sequence and expression in Escherichia coli of the
RT cephalosporin acylase gene of a Pseudomonas strain.";
RL Biochim. Biophys. Acta 1132:233-239(1992).
CC -!- FUNCTION: BESIDES THE CEPHALOSPORIN ACYLASE I ACTIVITY WHICH
CC CONVERTS GL-7ACA INTO 7-ACA; THIS ENZYME DISPLAYS SOME GAMMA
CC GLUTAMYLTRANSPEPTIDASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: 7-beta-(4-carboxybutanamido)-cephalosporanic
CC acid + H(2)O = 7-aminocephalosporanic acid + glutaric acid.
CC -!- SUBUNIT: DIMER OF TWO NON-IDENTICAL CHAINS PROCESSED FROM THE SAME
CC PRECURSOR.
CC -!- SIMILARITY: STRONG, TO OTHER GL-7ACA ACYLASES AND TO GGT ENZYMES.
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X69020; CAA48785.1; -  
 PIR; S27199; S27199.  
 MEROPS; T03.001; -  
 InterPro; IPR000101; Gslutrnspptdase.  
 Pfam; PF01019; G\_glu\_transpept; 1.  
 PRINTS; PR01210; GGTRANSPTASE.  
 TIGRFAMs; TIGR00066; g\_glu\_trans; 1.  
 PROSITE; PS00462; G\_GLU\_TRANSPEPTIDASE; 1.  
 Hydrolase; Antibiotic resistance; Zymogen; Transferase;  
 Acyltransferase.  
 INIT\_MET 0 0 BY SIMILARITY.  
 CHAIN 1 7366 ACYLASE ACY 1, LARGE SUBUNIT (BY  
 SIMILARITY).  
 CHAIN 367 557 ACYLASE ACY 1, SMALL SUBUNIT (BY  
 SIMILARITY).  
 SEQUENCE 557 AA; 57908 MW; 0C8386DACF7397A7 CRC64;

Query Match 7.6%; Score 96.5; DB 1; Length 557;

Best Local Similarity 23.0%; Pred. NO. 0.53;  
 Matches 59; Conservative 26; Mismatches 80; Indels 91; Gaps 10;

QY 49 NREPV-ALGVVPAALAAWREALA-----TAVERP----- 78  
 DB 116 DRENVGAKAVPGLGWCLEARFGLPLAEVLPQALGAERGFVTVYLSNCITDN 175  
 QY 79 -----DPGEP-----RRVPHDY-----HPPEGLFDRRCCLAAAM 110  
 DB 176 AADLARDPGLAAMLLPGQPLQPMRLQSDYASLKLIAGPEPALYGR---LGRALT 232  
 QY 111 DYPQDSIVIDMGT-----ATVIDLAGHFRGRILPGIAMSRLRGLHGTTALLPEVLNA 166  
 DB 233 DYNAANGLLIDQADLSYRIELREPIRGYRGYEIIGPPPTSSGVH--IAQMLNILEGY 290  
 QY 167 PAEMGLNDTSNAIOGVHFLFADALRGAITD-----FRQYSPQARI 207  
 DB 291 DIGALGFESTDA-----VHLLAEALKAFAADRAVATADPAFVKVPVARLIDKAYADERA 345  
 QY 208 LITGDAERWQPIAG 223  
 DB 346 LIAMEQAKSWTAGLSG 361

#### RESULT 4

LEU2\_RHIME STANDARD; PRT; 469 AA.  
 ID AC Q92L76; Q9EV53;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)  
 DE (isopropylmalate isomerase) (alpha-IPM isomerase) (IPMI).  
 GN LEUC OR R03206 OR SMC03823.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OC NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 RA Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;

\*Analysis of the chromosome sequence of the legume symbiont  
 Sinorhizobium meliloti strain 1021.\*;  
 Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).

[2]  
 RP SEQUENCE OF 22-469 FROM N.A.  
 RC STRAIN=GR4;  
 RA Sanjuan-Pinilla J.M., Munoz S., Olivares J., Sanjuan J.;  
 RT "The Sinorhizobium meliloti leuA gene is essential for symbiosis.";  
 RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC !- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate  
 CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.  
 CC !- CATALYTIC ACTIVITY: 3-isopropylmaleate -> 2-isopropylmaleate +  
 CC H(2)O.  
 CC !- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O -> 2-  
 CC isopropylmaleate.  
 CC !- PATHWAY: Leucine biosynthesis; second step.  
 CC !- SUBUNIT: Heterodimer of leuc and leud (By similarity).  
 CC !- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 1  
 CC SUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AL591793; CAC47785.1; -  
 EMBL; AJ296268; CAC14578.1; -  
 InterPro; IPR001030; Aconitase\_N.  
 InterPro; IPR004430; Leuc.  
 Pfam; PF00330; aconitase; 1.  
 ProDom; PD000511; Aconitase\_N; 1.  
 TIGRFAMs; TIGR00170; leuc; 1.  
 PROSITE; PS00450; ACONITASE\_1; 1.  
 PROSITE; PS01244; ACONITASE\_2; 1.  
 Leucine biosynthesis; lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
 FT METAL 350 350  
 FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT CONFLICT 31 31 R -> C (IN REF. 2).  
 FT CONFLICT 96 96 F -> L (IN REF. 2).  
 FT CONFLICT 270 270 T -> M (IN REF. 2).  
 FT CONFLICT 371 371 A -> S (IN REF. 2).  
 SQ SEQUENCE 469 AA; 50994 MW; 22A2F7D040645726 CRC64;

Query Match 7.4%; Score 94; DB 1; Length 469;

Best Local Similarity 25.9%; Pred. NO. 0.72;  
 Matches 58; Conservative 28; Mismatches 70; Indels 68; Gaps 13;

QY 5 AVGNTRTLAHTHDGVHFDVSV--ATSLPPTILOQGLTWLSAPNRPVAGGVVPA 62  
 DB 261 AVEYKTL--HTDEGAHYDRVVDLANLPPI-----VSWGSSP-EDVSVQGVVNP 310  
 QY 63 LAAREALAEVREPDGFFRRVPHDYHPPESGLFDRCCLLAAAMDYPGQDSIVDM 122  
 DB 311 DDIDQETKRTSKWR-----ALDYM-----LKP 333  
 QY 123 GTAIT--IDLLAGHFRGRI--LPGIAMSRLRGLHGTTALLPEV-VLNAPAEMLGNDTS 176  
 DB 334 GPKITDIAIDRVFIGSCTNGRIEDLRAVEV---EGRVAPTVSAMIVPGSLVKEQA 389  
 QY 177 NAIQGVHFLFADALRGAITDFRQYSPQARILITGGDAERWQPG 220  
 DB 390 EA--EGLDKIFEA-----GFDWRFPSCMCLAMD-DRLKPG 424

#### RESULT 5

CKX1\_ORYSA  
 ID CKX1\_ORYSA STANDARD; PRT; 532 AA.  
 AC Q9LDE6;  
 DT 15-JUN-2002 (Rel. 41, Created)









```

*RESULT 11
PEX6_HUMAN
ID PEX6_HUMAN STANDARD; PRT; 980 AA.
AC Q13608; Q99476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisome assembly factor-2 (PAF-2) (Peroxisomal-type ATPase 1)
DE (Peroxin-6).
GN PEX6 OR PXAAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96272151; PubMed=8670792;
RA Yahrus T., Braverman N., Dodi G., Kallish J.E., Morrell J.C.,
RA Moser H.W., Valle D., Gould S.J.;
RA "The peroxisome biogenesis disorder group 4 gene, PXAAL, encodes a
RA cytoplasmic ATPase required for stability of the PISI receptor.";
RL EMBO J. 15:2914-2923(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094178; PubMed=8940266;
RA Fukuda S., Shinozawa N., Suzuki Y., Zhang Z., Tomatsu S.,
RA Tsukamoto T., Hashiguchi N., Osumi T., Masuno M., Imaizumi K.,
RA Kuroki Y., Fujiki Y., Orii T., Kondo N.;
RA "Human peroxisome assembly factor-2 (PAF-2): a gene responsible for
RA group C peroxisome biogenesis disorder in humans.";
RL Am. J. Hum. Genet. 59:1210-1220(1996).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS ZS GLN-812 AND TRP-812.
RX MEDLINE=99335264; PubMed=10408779;
RA Zhang Z., Suzuki Y., Shinozawa N., Fukuda S., Imamura A.,
RA Tsukamoto T., Osumi T., Fujiki Y., Orii T., Wanders R.J.A.,
RA Barth P.G., Moser H.W., Paton B.C., Besley G.T., Kondo N.;
RA "Genomic structure and identification of 11 novel mutations of the
RA PEX6 'peroxisome assembly factor-2' gene in patients with peroxisome
RA biogenesis disorders.";
RL Hum. Mutat. 13:487-496(1999).
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR
CC STABILITY OF THE PISI RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DISEASE: DEFECTS IN PEX6 ARE THE CAUSE OF PEROXISOME BIOGENESIS
CC DISORDER GROUP 4 (PBD GROUP 4). PBD ARE A GROUP OF LETHAL
CC INHERITED DISEASES. THEY ARE CLINICALLY DIVIDED INTO FOUR GROUPS:
CC ZELLWEGER SYNDROME (ZS), NEONATAL ADRENOLKODYSTROPHY (NALD),
CC INFANTILE REFSUM DISEASE (IRD), AND CLASSICAL RHIZOMELIC
CC CHONDRODYSPLASIA PUNCTATA (RCDP). ZS IS THE MOST SEVERE PHENOTYPE
CC OF THE PBD AND IS CHARACTERIZED BY THE ABSENCE OF PEROXISOMES AND
CC BY MULTIPLE METABOLIC DEFECTS. PATIENTS WITH ZS HAVE SEVERE
CC NEUROLOGICAL ABNORMALITIES, DYSMORPHIC FACIAL FEATURES,
CC HEPATOMEGALY, AND MULTIPLE RENAL CYSTS; MOST DIE WITHIN 6 MONTHS
CC OF BIRTH. NALD AND IRD ARE SIMILAR TO ZS, BUT PROGRESSIVELY
CC MILDER, WITH SOME IRD PATIENTS SURVIVING BEYOND THE THIRD DECADE
CC OF LIFE. CLASSICAL RCDP, DISTINCT FROM THE ZS SPECTRUM, IS
CC ASSOCIATED WITH A CHARACTERISTIC FACIAL APPEARANCE, RHIZOMELIA,
CC ICTHYOSIS, AND SEVERE NEUROLOGICAL IMPAIRMENT.
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U56602; AAC50655.1; -
CC EMBL; D83703; BAA12069.1; -
CC EMBL; AF108098; AAF62564.1; -
CC EMBL; AF108095; AAF62564.1; JOINED.

```

```

DR EMBL; AF108096; AAF62564.1; JOINED.
DR EMBL; AF108097; AAF62564.1; JOINED.
DR Genew; HGNC:8859; PEX6.
DR MIN; 601498; -
DR MIM; 214100; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR Pfam; PF00004; AAA; 2.
DR SMART; SMO0382; AAA; 1.
DR PROSITE; PS00674; AAA; 1.
KW Peroxisome; ATP-binding; Repeat; Disease mutation.
FT NP_BIND 470 477 ATP (POTENTIAL).
FT NP_BIND 744 751 ATP (POTENTIAL).
FT VARIAT 812 812 R -> Q (IN ZS).
FT VARIAT 812 812 /FTIG-VAR_007918.
FT VARIAT 812 812 R -> W (IN ZS; ATYPICAL).
FT VARIAT 812 812 /FTIG-VAR_007919.
FT CONFLICT 77 77 S -> N (IN REF. 1).
SQ SEQUENCE 980 AA; 104060 MW; 0EC1C2A75CE0038F CRC64;
Query Match 7.0%; Score 88.5; DB 1; Length 980;
Best Local Similarity 21.3%; Pred. No. 5.2;
Matches 54; Conservative 27; Mismatches 89; Indels 83; Gaps 10;
QY 25 TSLPTEILQOPLTWSAPNREPVALGVV-----PAAALANREALATAEVREPDGFFR 84
DB 14 TETPLAVLLPPGPM-----PAALGLVLALRPAESPAGPALLVALEPGDAG--- 63
QY 85 RAVPHDYHPPESLGRDRCCLLAAAMDYQDSIVIDMTAITIDLLAGHFRGRL-- 142
DB 64 -----TEEQG-----PGPQLLVSRALLRLALLGSGANVVARVRP 100
QY 143 PGIAMSLRGLHEGTALLPEV-----VLNAPAEMLGNDTSNAIOAGVIHLFADALRG 193
DB 101 PALGWAALLGTSGLPGSLGRVGLVRRGETLPVPGPRV-LVETRALQG-----LLGPTRL 155
QY 194 AITDPRQVSPQARILITGDDAERWP-----GIAGSLY 226
DB 156 AVTELR---GRARICPESGDSRPPPPVSVFVSGTVRRVLRQGLVGGTSLGVSRSKL 212
QY 227 QPHLLRGFYLMV 239
DB 213 RGLGLFQGEVWV 225
RESULT 12
ARGD_MYCTU STANDARD; PRT; 400 AA.
AC P94990;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetylornithine aminotransferase (EC 2.6.1.11) (ACOAAT).
GN ARGD OR RV1655 OR MV1693 OR MTCY06H11.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";

```



RC STRAIN=SP7;  
 RX MEDLINE=97094331; PubMed=8939798;  
 RA de Troch P., Dosselaere F., Keijers V., de Wilde P.,  
 RA Vanderleyden J.;  
 RT "Isolation and characterization of the Azospirillum brasilense  
 RT trpE(g) gene, encoding anthranilate synthase.";  
 RL Curr. Microbiol. 34:27-32(1997).  
 CC -1- CATALYTIC ACTIVITY: Chorismate + L-glutamine -> anthranilate +  
 CC pyruvate + L-glutamate.  
 CC -1- PATHWAY: Tryptophan biosynthesis; first step.  
 CC -1- SIMILARITY: ITS FIRST TWO-THIRDS SHOWS SIMILARITY TO THE COMPONENT  
 CC I OF ANTHRANILATE SYNTHASE FROM OTHER BACTERIA, WHEREAS THE FINAL  
 CC ONE-THIRD RESEMBLES THE COMPONENT II OF THE SAME BACTERIA.  
 CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U44127; AAC45141.1;  
 CC HSPB: P00897; I17S  
 CC InterPro: IPR002385; Anth\_synthII.  
 CC InterPro: IPR000350; Chorismate\_bind.  
 CC InterPro: IPR000991; GATase\_1.  
 CC InterPro: IPR001789; Response\_reg.  
 CC Pfam: PF00117; GATase\_1.  
 CC Pfam: PF00425; chorismate\_bind; 1.  
 CC PRINTS: PRO0095; ANTSNTHASEI.  
 CC PRINTS: PRO0097; ANTSNTHASEII.  
 CC PRINTS: PRO0096; GATase.  
 CC ProDom: PD000779; Chorismate\_bind; 1.  
 CC TIGRFAMS: TIGR00566; trpG\_papa; 1.  
 CC PROSITE: PS00442; GATASE\_TYPE\_1; 1.  
 CC Tryptophan biosynthesis; Lyase; Glutamine amidotransferase.  
 FT ACT\_SITE 610 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 699 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 701 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 732 AA; 78041 MW; 4195A53D381B4B7A CRC64;  
 -----  
 Query Match 6.88; Score 86; DB 1; Length 732;  
 Best Local Similarity 26.78; Pred. No. 6.1;  
 Matches 56; Conservative 18; Mismatches 62; Indels 74; Gaps 12;  
 Y 33 PTEILOQGLTWSAPNRPVAL---GV-----VPAALAAWREALAT 72  
 D 3 PADLLASPDLL-----EPLRFQTRGVTVTRATALDPTALDPVIDALDRRGLLS 55  
 Q 73 AEVRPDPGFRRRAVPHDYHPPEISGLFDRRCCLLAAMADYPCQDSIVIDMTATIDLLA 132  
 D 56 SGVEA---PGRYR-----HALGTFDPAVALTA-----RGRTLRDALN 91  
 Q 133 GGHRRGRRLPGIAMSRLGLHEGTALLPEVVLN-----APAEMLGNDTSN-----AQQA 181  
 D 92 G---RGQVLLPAVAELRGL--EALAGLEAPSRVTASSAPAPLPGEERSQPSVFLR 147  
 Q 182 GVHLEA---DALRGATIDR-----QVSP 203  
 D 148 AVLDLFAAPDDPLLGAGAYDLAFQFEP 177  
 RESULT 15  
 FTSH\_MYCTU  
 ID FTSH\_MYCTU STANDARD; PRT; 760 AA.  
 AC P96942;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cell division protein ftsh homolog (EC 3.4.24.-).  
 GN FTSH OR RV3610C OR MT3714 OR MTCY07H7B.12.

OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 203-323 FROM N.A.  
 RC STRAIN=ATCC 201 / H37Ra;  
 RX MEDLINE=98391006; PubMed=9729123;  
 RA Anilkumar G., Chauhan M.M., Ajitkumar P.;  
 RT "Cloning and expression of the gene coding for FtsH protease from  
 RT Mycobacterium tuberculosis H37Rv.";  
 RL Gene 214:7-11(1998).  
 CC -1- FUNCTION: SEEMS TO ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE  
 CC (BY SIMILARITY).  
 CC -1- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M41.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z95557; CAB08956.1;  
 CC EMBL: AE007171; AAK48073.1;  
 CC EMBL: Z83338; CAB05953.1;  
 CC MEROPS: M41.UPW;  
 CC TIGR: MT3714;  
 CC TubercuList; RV3610C;  
 CC InterPro: IPR003593; AAA\_ATPase.  
 CC InterPro: IPR003959; AAA\_ATPase\_cent.  
 CC InterPro: IPR003960; AAA\_sub.  
 CC InterPro: IPR000642; Peptidase\_M41.  
 CC Pfam: PF00004; AAA; 1.  
 CC SMART: SM00382; Peptidase\_M41; 1.  
 CC Pfam: PF01434; AAA; 1.  
 CC TIGRFAMS: TIGR01241; FtsH\_fam; 1.  
 CC PROSITE: PS00674; AAA; 1.  
 CC Cell division; ATP-binding; Transmembrane; Hydrolase; Metalloprotease;  
 KW zinc; Complete proteome.  
 FT DOMAIN 1 5 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT DOMAIN 27 110 EXTRACELLULAR (POTENTIAL).

```
FT TRANSMEM 111 131 POTENTIAL.
FT DOMAIN 132 760 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 203 210 ATP (POTENTIAL).
FT METAL 425 425 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 426 426 BY SIMILARITY.
FT METAL 429 429 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 390 390 V -> I (IN REF. 2).
SQ SEQUENCE 760 AA; 81985 MW; AFEF8DBF10E7C6CF CRC64;

Query Match
Best Local Similarity 21.5%; Score 85; DB 1; Length 760;
Matches 51; Conservative 27; Mismatches 69; Indels 90; Gaps 10;

Qy 41 GLTWLSAPNRPVVALGGVVPALAAWREALATAEVREPPGFFRRAVPHDYHPPESLGFD 100
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
301 GVILIAATNRDPI-----LDPALLR-----PGRFDRQIP--VSNPDLAG-- 337

Qy 101 RRCCI-----LAAANDYPQDSIVDM-----GTAIT----- 127
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
338 RRAVLRVHSGKPKMAADADLDGLAKRTVGMTGADLANVINEAALLTARENGVITGPAL 397

Qy 128 --IDLAGGHFRGRILFGIAMSRLRHEGTALL-----PEV-----V 163
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
398 EAVDRVIGGPRKGRRIIEQEKKITAYHEGGHTLAAWAMPDIEPIYKVITLARGTGGHA 457

Qy 164 LNAPEMLGNDTSNIAIQGVTH-----LFDALRGATDFRQYSPQARILIT 210
Db : : | | : : | | : : | | : : | | : : | | : : | | : : | | : : | |
458 VAVPEEDKGLRTRSEMIQLVFAMGGRAAEELVFREPTTGAVSDIEQATKIARSMVT 514
```

Search completed: June 24, 2003, 22:11:58  
Job time : 5.84794 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 27.2799 Seconds  
(without alignments)  
1177.181 Million cell updates/sec

Title: US-09-813-453A-63

Perfect score: 1264

Sequence: 1 MFIAGVNRFTLLAHTHDGV.....AGSLYQPHLLRGFYLIWIRG 241

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1264	100.0	241	AAU91179	Pantothenate kinas
2	229.5	18.2	246	AAU91156	Thermotoga maritim
3	227	18.0	257	AAU91174	Pantothenate kinas
4	227	18.0	260	AAU91173	Pantothenate kinas
5	215	17.0	262	AAU91155	Deinococcus radiop
6	203	16.1	258	AAU01243	B. subtilis novel
7	203	16.1	258	AAU91149	Bacillus subtilis
8	190.5	15.1	262	AAU91170	Pantothenate kinas
9	189	15.0	219	AAU91176	Pantothenate kinas
10	185	14.6	259	23 ABB47661	Listeria monocytog

11	184	14.6	265	23	AAU91151	Streptomyces coeli
12	183.5	14.5	244	23	AAU91168	Pantothenate kinas
13	181	14.3	232	23	AAU91163	Pantothenate kinas
14	180.5	14.3	212	23	AAU91177	Pantothenate kinas
15	179.5	14.2	254	23	AAU91171	Pantothenate kinas
16	178.5	14.1	258	23	AAU91172	Pantothenate kinas
17	174	13.8	257	23	AAU91160	Synechocystis pant
18	172.5	13.6	255	23	AAU91154	Geobacter sulfure
19	170.5	13.5	273	23	AAU91157	Treponema pallidum
20	169	13.4	250	23	AAU91150	Clostridium acetob
21	168.5	13.3	258	23	AAU91153	Rhodobacter capsul
22	164.5	13.0	229	23	AAU91159	Aquifex aeolicus p
23	159	12.6	272	22	AAU91152	Mycobacterium tube
24	159	12.6	272	23	AAU91152	Mycobacterium tube
25	154.5	12.2	267	23	AAU91162	Bordetella pertussis
26	145	11.5	256	23	AAU91175	Pantothenate kinas
27	144.5	11.4	223	19	AAU98422	H. pylori GHPO 344
28	144.5	11.4	223	23	AAU91161	Helicobacter pylor
29	144.5	11.4	223	23	AAU91181	Pantothenate kinas
30	144.5	11.4	248	23	AAU91164	Pantothenate kinas
31	143.5	11.4	242	23	AAU91180	Pantothenate kinas
32	136	10.8	262	23	AAU91158	Borrelia burgdorfe
33	124	9.8	209	23	AAU91165	Pantothenate kinas
34	123.5	9.8	455	20	AAU93617	Neisseria gonorrhoe
35	123.5	9.8	455	21	AAU91167	Neisseria gonorrhoe
36	123.5	9.8	460	23	AAU91167	Pantothenate kinas
37	123.5	9.8	592	21	AAU93618	Neisseria gonorrhoe
38	123.5	9.8	592	21	AAU94911	Neisseria meningit
39	116	9.2	389	21	AAU94909	Neisseria meningit
40	115.5	9.1	592	20	AAU93615	Neisseria meningit
41	115.5	9.1	592	21	AAU94912	Neisseria meningit
42	115.5	9.1	592	23	AAU91169	Pantothenate kinas
43	112.5	8.9	455	21	AAU94910	Neisseria meningit
44	112.5	8.9	592	20	AAU93616	Neisseria meningit
45	112.5	8.9	592	21	AAU94913	Neisseria meningit

#### ALIGNMENTS

##### RESULT 1

AAU91179  
ID AAU91179 standard; Protein; 241 AA.

AC AAU91179;

DT 05-JUN-2002 (first entry)

DE Pantothenate kinase (Coax) #17.

XX Pantothenate kinase; Coax; antitibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Thiobacillus ferrooxidans.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX N-PSDB; ABR54200.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT pantothenate kinase comprising pantothenate kinase (Coax) protein  
PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein  
 PS Claim 10; Page 109-110; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 241 AA;

Query Match 100.0%; Score 1264; DB 23; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 9e-132;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 MIFTAVGNTRTLAAHTHDGVHFDVSVSATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP 60  
 1 MIFTAVGNTRTLAAHTHDGVHFDVSVSATSLPPTTEILQQPGLTWLSAPNREPVALGGVVP 60  
 61 AALAAAREALATAEVRPDPGFFRRVAPHDYHPPESLGDFRCCLLAAAMDYPGQDSIVI 120  
 61 AALAAAREALATAEVRPDPGFFRRVAPHDYHPPESLGDFRCCLLAAAMDYPGQDSIVI 120  
 121 DMGTAITDLLAGHFRGGRILPGIAMSRLGHEGTALLPEVNLNAPAEMLGNDTSNAIQ 180  
 121 DMGTAITDLLAGHFRGGRILPGIAMSRLGHEGTALLPEVNLNAPAEMLGNDTSNAIQ 180  
 181 AGVHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRGYLWIR 240  
 181 AGVHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRGYLWIR 240  
 241 G 241  
 241 G 241

RESULT 2  
 AAU91156  
 ID AAU91156 standard; Protein; 246 AA.  
 XX  
 AC AAU91156;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 Thermotoga maritima pantothenate kinase Coax.  
 ..A  
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Thermotoga maritima.  
 OS  
 XX WO200216601-A2.  
 PN  
 XX 28-FEB-2002.  
 PD  
 XX 24-AUG-2001; 2001WO-US26531.  
 PF  
 XX 24-AUG-2000; 2000US-227860P.  
 PR  
 XX 20-MAR-2001; 2001US-0813453.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX  
 PI Yocum RR, Patterson TA;  
 XX  
 DR WPI; 2002-269358/31.  
 XX  
 PT Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 10; Page 74-75; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 246 AA;

Query Match 18.2%; Score 229.5; DB 23; Length 246;  
 Best Local Similarity 28.2%; Pred. No. 7e-17;  
 Matches 71; Conservative 40; Mismatches 110; Indels 31; Gaps 5;  
 2 IFIAGVNTRTLAAHTHDGVHFDVSVSATSLPPTTEILQQPGLTWLSAPNREPVALG--GV 58  
 3 LLVDVGNTHSVFSTEDKTFRRWLSTGVFTQTEDELFSHLHPLLDGAMREIKGIGVASV 62  
 59 VPAALAA-----WREALATAEVRPDPGFFRRVAPHDYHPPESLGDFRCC 104  
 63 VPTQNTVIERFSQYFHISPIWVAK-----NGCVKNVKNPSEVGADRVAN 109  
 105 LLAAAMDYPGQDSIVIDMGTAITDILLAGHFRGGRILPGIAMSRLGHEGTALLPEVL 164  
 110 VVAFVKEY--GKNGIIDIIMGTATVDLVNVSYGGAIFLPGFTMMVHSLFRGTAKPLVEV 168  
 165 NAPAEMLGNDTSNAIQAGVHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGS 224  
 169 KPADFVVGKDFEENIRLGLVNGSVTALGIIIGRIKEIVYGDLPVLTGGQSKIYKDMIKHE 228  
 225 LYQPHLLLRGYF 236  
 229 IFDEDLTKGVY 240

RESULT 3  
 AAU91174  
 ID AAU91174 standard; Protein; 257 AA.  
 XX  
 AC AAU91174;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 Pantothenate kinase (Coax) #12.  
 DE  
 XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 KW  
 OS Chlorobium tepidum.  
 XX  
 PN WO200216601-A2.  
 XX  
 PD 28-FEB-2002.  
 PF  
 XX 24-AUG-2001; 2001WO-US26531.  
 PR  
 XX 24-AUG-2000; 2000US-227860P.  
 PR  
 XX 20-MAR-2001; 2001US-0813453.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX  
 PI Yocum RR, Patterson TA;  
 XX  
 DR WPI; 2002-269358/31.  
 DR N-PSDB; ABK54195.

```

XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 10; Page 103-104; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
XX Sequence 257 AA;
XX
Query Match 18.0%; Score 227; DB 23; Length 257;
Best Local Similarity 32.1%; Pred. No. 1.4e-16;
Matches 84; Conservative 43; Mismatches 101; Indels 34; Gaps 12;
QY 2 IFIATVGNTRLLAHTHDGVFDSV-SVATSLPPTTEILQQPGLTWLSAPNREP--VALGGV 58
Db : : ||| ||| : ||| : : : : : : : : : : : : : : : : : : : :
3 LVYDIGNTSTTLA-IFTGDEEPSVESVPSALFADSDTMREVFNGMARKHGEPOAICSV 61
QY 59 VPAALA---AWREALATAEV-----REPDGCFERRAVPHDHPHPSLGFDRCCLLAA 108
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
62 VPSRTAVGSALLSLSFVPLTTCCKLRFP-----FRL----DYATPTFGADRLALCAWS 113
QY 109 AMDYPGQDSVIDMGTAITDILL-AGGHFRGGRILPGIAMSRLGLHEGTALLPEVVLNAP 167
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
114 RHLFSEKPVIAVDIGTAITDVLDTVGNRGGLIMPCIDMMAGALHSRTAQLPQVRIDRP 173
QY 168 AEMLGNDTSNAIQGV-----IHLFADALRG-AITDFRQYSPQARILITGGDAERWQP 219
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
174 ESLIGRSTTECIKSGVFWGVVKGIGGLVDALRGDLVRDFGEST--VEVITVGGNSRIIVP 231
QY 220 GIAG-SLYQPHLLLRGYLWIR 240
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
232 EIGPVSVIDELAVLRGSDLLLR 253
XX
RESULT 4
U91173
AAU91173 standard; Protein; 260 AA.
XX
AC AAU91173;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #11.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Caulobacter crescentus.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
XX
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX Yocum RR, Patterson TA;
XX

```

```

DR WPI; 2002-269358/31.
DR N-PSDB; ABK54194.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
XX Claim 10; Page 102-103; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX
XX Sequence 260 AA;
XX
Query Match 18.0%; Score 227; DB 23; Length 260;
Best Local Similarity 29.0%; Pred. No. 1.4e-16;
Matches 75; Conservative 43; Mismatches 103; Indels 38; Gaps 12;
QY 1 MIFTAVGNTRLLAHTHDG----VHFDSVSVATSLPP-----TEILOQPGGLTWLSAPNR 50
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
2 LLATEQGNNTMFA-IHDGASVAVQWRSATESTRTADEVVWLSQLLSMQLGFRAL--- 57
QY 51 EPVALGGVVPALAAWR-----EALATAEVREPDPGFFRRAVPHDHPHPSLGFDR 101
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
58 DAVTISVVQPSIFNRLNLSRYFVPLVIGENA-----KLGIDVRIKPESEAGADR 110
QY 102 RCCLLAAAMDYPGQDSVIDMGTAITDILLAG-GHFRGGRILPGIAMSRLGLHEGTALLP 160
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
111 LVNAIGAAMVTPG-PLVVIDSGTATFDIVAADGAFEGGIAPGINLSWQALHEAAKLP 169
QY 161 EYVLNAPA--EMLGNDTSNAIQGVYIHLFPADALRGAITDFR-QYSPQARILITGGDAERW 217
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
170 RIAIQRPAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTVIATGVSALF 229
QY 218 QPGTAGSL--YQPHLLLRG 234
Db : : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
230 E-GATDSIDHFDSDLTIRG 247
XX
RESULT 5
AAU91155
ID AAU91155 standard; Protein; 262 AA.
XX
AC AAU91155;
XX
DT 05-JUN-2002 (first entry)
XX
DE Deinococcus radiopugnans pantothenate kinase Coax.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Deinococcus radiopugnans.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
XX
XX 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
XX

```







PN WO200216601-A2.  
 XX 28-FEB-2002.  
 PD 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX Yocum RR, Patterson TA;  
 PI WPI; 2002-269358/31.  
 DR N-PSDB; ABK54197.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX  
 --s Claim 10; Page 106-107; 128pp; English.  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 SQ Sequence 219 AA;  
 Query Match 15.0%; Score 189; DB 23; Length 219;  
 Best Local Similarity 28.6%; Pred. No. 1.8e-12;  
 Matches 66; Conservative 37; Mismatches 78; Indels 50; Gaps 9;  
 QY 1 MIFAVGNT-----RTLLAH---THDGVHFDVSVATSLPP-----TEILOQPGTWL 45  
 Db LVAVDIGNTSVNIGIFEGEKLLANWHLG-----SVAQRMADEVASLLGLQHAG---I 55  
 QY 46 SAPNRPVALGGVVPVPLTTTFFEVFKS-----YFKAAPLVGAGIKSGVKVRMDNP 94  
 Db HPEELNRVIMCSVVPPLTTTFFEVFKS-----YFKAAPLVGAGIKSGVKVRMDNP 106  
 QY 95 ESLGFRRCCLLAAMDYPGQDSIVIDMGTAITDILL-AGHFRGGRILPGIAMSRLGLH 153  
 107 REVGADRIYNNAARVLYPGA-CIIVDMGTATTFDLISEGAYIGGAIATSAQAIA 165  
 Y 154 ECTALLPEVVLNAPAEMLGNDTSNAIQAGV----IHLFADALRGATDFRQ 200  
 Db 166 EKTSLPKTEIIRPAKVIQNTVSAMQSGIYFGYIGIVEELVRRIQELQ 216  
 RESULT 10  
 ID ABB47661  
 XX ABB47661 standard; Protein; 259 AA.  
 AC ABB47661;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #365.  
 XX  
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN WO200177335-A2.  
 XX

PD 18-OCT-2001.  
 XX 11-APR-2001; 2001WO-FR01118.  
 XX 11-APR-2000; 2000FR-0004629.  
 XX (INSP ) INST PASTEUR.  
 PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 DR Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 XX and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -  
 PT  
 PS Claim 6; SEQ ID No 366; 192pp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 259 AA;  
 Query Match 14.6%; Score 185; DB 23; Length 259;  
 Best Local Similarity 27.0%; Pred. No. 6.6e-12;  
 Matches 72; Conservative 43; Mismatches 108; Indels 44; Gaps 12;  
 QY 1 MIFAVGNT-----RTLLAH---THDGVHFDVSVATSLPTEILOQPGTWSAP 48  
 Db 2 ILVDVGNCTGTVGVYKQKLLKHWRMTDR-HRTSDELGMTV-----LNFFSYA 50  
 QY 49 NREP-----VALGGVVPVPAALAAWREALATAEVR---BP---DPCGFERRAVPHDYPPES 96  
 Db 51 NLTPSDIOGIITISSVVPPI-----HAMETMCRYNIRPLIVPG-ITGLNMLKVDNPRE 105  
 QY 97 LGFDRRCCLLAAMDYPGQDSIVIDMGTAITDILL-AGHFRGGRILPGIAMSRLGLHEG 155  
 Db 106 IGSRIYNNAVAASEEY-GTPVIVVDFTATFCYIDESGVYGGGAIPGIMSTEALYNR 164  
 QY 156 TALLPEVVLNAPAEMLGNDTSNAIQAGVHIHFLFADALRGATDFROYSPQARILITGGDAE 215  
 Db 165 AAKLPVRDIAESSQIGKSTVSSMQAGIFYGVQCEGIIAEMKQSNASPVVATGGLA 224  
 QY 216 RW--QPGTAGSLYOPHLLLRGFWLWR 240  
 Db 225 RMITEKSSAVDILDFFLTGLKLELLYR 251  
 RESULT 11  
 ANU91151

AAU91151 standard; Protein; 265 AA.  
 AAU91151;  
 05-JUN-2002 (first entry)  
 Streptomyces coelicolor pantothenate kinase Coax.  
 Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 Streptomyces coelicolor.  
 WO200216601-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US26531.  
 24-AUG-2000; 2000US-227860P.  
 20-MAR-2001; 2001US-0813453.  
 (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 Yocum RR, Patterson TA;  
 WPI; 2002-269358/31.  
 Identifying potential antibiotic or antimicrobial agent, comprises  
 contacting composition comprising pantothenate kinase (Coax) protein  
 with test compound and identifying inhibitor of the Coax protein -  
 Claim 10; Page 69-70; 128pp; English.  
 The invention describes assays for identifying a (potential) antibiotic  
 comprising contacting an assay composition comprising a pantothenate  
 kinase (Coax) protein with a test compound, and determining the ability  
 of the test compound to inhibit the activity of the Coax protein, an  
 essential enzyme for the production of coenzyme A. Coax protein is a  
 valuable target for identifying bactericidal compounds. Coax modulating  
 agents can be used in an infectious animal model to determine the  
 efficacy, toxicity, or side effects of treatment with such an agent. This  
 is the amino acid sequence of a pantothenate kinase (Coax) protein  
 described in the invention.

Query Match 14.6%; Score 184; DB 23; Length 265;  
 Best Local Similarity 28.8%; Pred. No. 8.8e-12;  
 Matches 76; Conservative 39; Mismatches 113; Indels 36; Gaps 11;  
 QY 1 MIFTAVGNTRTLTAHHDG-----VHFDVSVAATSLPPTILQ-----QPGITWLSA 47  
 DB 2 LLRIDVGNTHVLG-LFDGEDIVHWRISTDSRTADEL--AVILQGLMGHPILGDELG 58  
 QY 48 PNREPAVGGVVPAALAAWREALATAEVREP-----DPGFFRRAVPHDYHPPEISLGFDRRC 103  
 DB 59 DGIDGIAICATVPSPVHLEHREVTTRYGDPVAVLVEPG-VKTGVPILTDHPKEVGADRII 117  
 QY 104 CLLAAAMDYPCQDSIVIDMGTAITIDLL-AGGHFRGGRIILPGIAMSRLGLHECTALLPEV 162  
 DB 118 NANAARELY-GGPAIVVDFTFTTDAVSARGEYVIGVIAEGISVEALGVKGAQLRKI 176  
 QY 163 VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAIT----DFRQYSPQARILITGGDAERWQ 218  
 DB 177 EVAPRSVIGKNTVEAMQSGIVYGFAGQVDGVNVMARELADDPDDVTIVATGGLA----- 232  
 QY 219 PGIAAGSL-----YQHLLLRGFFYL 237  
 DB 233 PMVLGESSVIDEHEPWLTLMLRL 256

RESULT 12

AAU91168  
 ID AAU91168 standard; Protein; 244 AA.  
 AC AAU91168;  
 DT 05-JUN-2002 (first entry)  
 DE Pantothenate kinase (Coax) #6.  
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 OS Porphyromonas gingivalis.  
 XX WO200216601-A2.  
 PN 28-FEB-2002.  
 PD 24-AUG-2001; 2001WO-US26531.  
 PF 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 PA Yocum RR, Patterson TA;  
 PI WPI; 2002-269358/31.  
 DR N-PSDB; ASK54189.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 PS Claim 8; Page 94-95; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.

Query Match 14.5%; Score 183.5; DB 23; Length 244;  
 Best Local Similarity 35.4%; Pred. No. 8.9e-12;  
 Matches 51; Conservative 23; Mismatches 65; Indels 5; Gaps 3;  
 QY 96 SLGFFRRCLLAAAMDYPCQDSIVIDMGTAITID-LLAGGHFRGGRIILPGIAMSRLGLHE 154  
 DB 92 TLGADRLAAVGAHSLYPNTELLVDAGTAITYERSAEGIVLGNISPGHLRFRKALHL 151  
 QY 155 GTALPEV---VLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITG 211  
 DB 152 FTGRPLDIPSGISPKIAEYGSSTEARITAGVIHGLAGEIDRYIDDLHAKGRSAVILTG 211  
 QY 212 GDAERWQPGI-AGSLYQPHLLLRG 234  
 DB 212 GDANYLARIIRSGILTHPDLVLG 235

RESULT 13

AAU91163  
 ID AAU91163 standard; Protein; 233 AA.  
 AC AAU91163;  
 XX 05-JUN-2002 (first entry)

```

XX DE Pantothenate kinase (Coax) #1.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Bacillus subtilis.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX N-PSDB; ABK54169.
XX WPI; 2002-269358/31.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Disclosure; Page 81-82; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 233 AA;
Query Match 14.3%; Score 181; DB 23; Length 233;
Best Local Similarity 27.9%; Pred. No. 1.6e-11;
Matches 61; Conservative 34; Mismatches 102; Indels 22; Gaps 7;
QY 1 MIFAVGNTRTLLAHTHDGVHDSVSATSLPPE-----ILQQPGLTWLSAPNRE 51
2 LLVIDGNTVTLGVYHDGKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMF---EQID 58
52 FVALGGVVPAAALAAWREALATAEAREPD---PGFFRAVPHDYHPPESLGFDRRCCLAA 108
59 GIISSVVPIMFALEMCYKYPHIEPQIVGPG-MKTGLNLIKNDPNKEVGADRIYNVAA 117
109 AMDPGODSVIDMGTAIT---IDLAGHFRGRILPGIAMSRLGHECTALLPEVYN 165
118 IHLY-GNPLIVDFGTATTTCYID--ENKQYMGAIAPGITISTEALYSRAAKLPRIET 174
166 APAMELGNDSNAIQAGVIHLFDALRGATDTRQYSPQ 204
175 RPDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMKQAKQ 213

```

```

RESULT 14
AAU91177
ID AAU91177 standard; Protein; 212 AA.
XX AC
XX AC AAU91177;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #15.

```

```

XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX OS Desulfovibrio vulgaris.
XX PN WO200216601-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26531.
XX PR 24-AUG-2000; 2000US-227860P.
XX PR 20-MAR-2001; 2001US-0813453.
XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX PI Yocum RR, Patterson TA;
XX N-PSDB; ABK54198.
XX WPI; 2002-269358/31.
XX PT Identifying potential antibiotic or antimicrobial agent, comprises
XX PT contacting composition comprising pantothenate kinase (Coax) protein
XX PT with test compound and identifying inhibitor of the Coax protein -
XX PS Claim 10; Page 107-108; 128pp; English.
XX CC The invention describes assays for identifying a (potential) antibiotic
XX CC comprising contacting an assay composition comprising a pantothenate
XX CC kinase (Coax) protein with a test compound, and determining the ability
XX CC of the test compound to inhibit the activity of the Coax protein, an
XX CC essential enzyme for the production of coenzyme A. Coax protein is a
XX CC valuable target for identifying bactericidal compounds. Coax modulating
XX CC agents can be used in an infectious animal model to determine the
XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
XX CC described in the invention.
XX SQ Sequence 212 AA;
Query Match 14.3%; Score 180.5; DB 23; Length 212;
Best Local Similarity 29.4%; Pred. No. 1.5e-11;
Matches 63; Conservative 31; Mismatches 83; Indels 37; Gaps 9;
QY 6 VGNTRTLIAHTHDGVHDSVSATSLPPEITLQQPGLTWLSAPNR--EPVALGGVVPAAAL 63
10 IGNTNVKI-----GLAVETAVLTSTVLPD---PGQTTDSIGLRLLEVLRHAGLGPADV 60
64 AAWREALATAEVRPDPGFFRR-----AVPHD---YHPPESLGDFDRRC 104
61 GA---CVASSVVPVGNP-LIRRACERYLYRKLIFAPGDTAIPLDNRYERPAEVGADRLVA 116
105 LLAAMDPGODSIV-IDMGTAITIDLAGHFRGRILPGIAMSRLGHECTALLPEVY 163
117 AYAAARLYPGPRSLVSDVFGTATTFDCEGGAYLGGICPGVLSSAGALSSRTAKLPRI 176
164 LNAPEAE--MLGNDTSNAIQAGVIHLFDALRGAI 195
177 LEVEEDSPVIGRSTTSLNHGFIQFGFAAMTEGVL 210

```

```

RESULT 15
AAU91171
ID AAU91171 standard; Protein; 254 AA.
XX AC
XX AC AAU91171;
XX DT 05-JUN-2002 (first entry)
XX DE Pantothenate kinase (Coax) #9.
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 12.4468 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-10

Perfect score: 1379

Sequence: 1 MLLIDVGNHVVFGQENG.....ARLVPTSLLPATVSGSGSN 273

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

atal number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1379	100.0	273	9	US-09-813-453A-10
2	371.5	26.9	255	9	US-09-813-453A-7
3	354	25.7	256	9	US-09-813-453A-55
4	342.5	24.8	258	9	US-09-813-453A-49
5	342.5	24.8	262	9	US-09-813-453A-45
6	337.5	24.5	260	9	US-09-813-453A-51
7	327.5	23.7	258	9	US-09-813-453A-2
8	320.5	23.2	254	9	US-09-813-453A-47
9	310	22.5	219	9	US-09-813-453A-57
10	293.5	21.3	258	9	US-09-813-453A-6
11	291.5	21.1	265	9	US-09-813-453A-4
12	288	20.9	233	9	US-09-813-453A-17
13	285	20.7	246	9	US-09-813-453A-9
14	282.5	20.5	250	9	US-09-813-453A-3
15	254	18.4	272	9	US-09-712-363-276
16	254	18.4	272	9	US-09-813-453A-5
17	248	18.0	262	9	US-09-813-453A-8
18	197.5	14.3	257	9	US-09-813-453A-53
19	192.5	14.0	212	9	US-09-813-453A-59

20	172.5	12.5	244	9	US-09-813-453A-41	Sequence 41, Appl
21	170.5	12.4	241	9	US-09-813-453A-63	Sequence 63, Appl
22	170	12.3	262	9	US-09-813-453A-11	Sequence 11, Appl
23	147.5	10.7	592	9	US-09-813-453A-22	Sequence 22, Appl
24	147.5	10.7	592	9	US-09-813-453A-43	Sequence 43, Appl
25	146.5	10.6	460	9	US-09-813-453A-39	Sequence 39, Appl
26	136.5	9.9	257	9	US-09-813-453A-13	Sequence 13, Appl
27	136	9.9	248	9	US-09-813-453A-20	Sequence 20, Appl
28	132.5	9.6	249	9	US-09-813-453A-70	Sequence 70, Appl
29	122.5	8.9	229	9	US-09-813-453A-12	Sequence 12, Appl
30	118	8.6	249	9	US-09-813-453A-61	Sequence 61, Appl
31	92.5	6.7	351	9	US-09-991-211-11	Sequence 11, Appl
32	92	6.7	604	10	US-09-758-269-14	Sequence 14, Appl
33	92	6.7	4053	9	US-10-123-155-351	Sequence 351, App
34	88	6.4	242	9	US-09-813-453A-65	Sequence 65, Appl
35	88	6.4	268	10	US-09-815-242-11977	Sequence 11977, A
36	87.5	6.3	3519	9	US-09-808-880-4	Sequence 4, Appl
37	85	6.2	280	9	US-09-712-363-278	Sequence 278, App
38	85	6.2	530	9	US-09-738-626-4188	Sequence 4188, Ap
39	82.5	6.0	390	9	US-09-712-363-192	Sequence 192, App
40	82	5.9	922	9	US-09-855-754-6	Sequence 6, Appl
41	82	5.9	922	9	US-10-227-353-6	Sequence 2, Appl
42	82	5.9	4150	9	US-09-808-880-2	Sequence 2, Appl
43	81.5	5.9	223	9	US-09-895-913A-74	Sequence 74, Appl
44	81.5	5.9	223	9	US-09-813-453A-14	Sequence 14, Appl
45	81.5	5.9	223	9	US-09-813-453A-67	Sequence 67, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-813-453A-10  
; Sequence 10, Application US/09813453A

; Patent No. US20020168681A1

; GENERAL INFORMATION:

; APPLICANT: Yocum, R. Rogers

; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; FILE REFERENCE: OGZ-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; PRIOR FILING DATE: 2000-03-20

; PRIOR APPLICATION NUMBER: US 60/227,860

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: 09/667,569

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 10

; LENGTH: 273

; TYPE: PRT

; ORGANISM: Treponema pallidum

US-09-813-453A-10

Query Match 100.0%; Score 1379; DB 9; Length 273;  
Best Local Similarity 100.0%; Pred. No. 1.8e-121;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLLIDVGNHVVFGQENGRCVCFRLPLAPDARKTQDEYSLLIHALCERAGVGRASL	60
Db	1	MLLIDVGNHVVFGQENGRCVCFRLPLAPDARKTQDEYSLLIHALCERAGVGRASL	60
Qy	61	RDAFTSSVVPVLTKTADAVQISGVQPVVFGWYAEHLPVRTPEPVRAEIGTDLVANAV	120
Db	61	RDAFTSSVVPVLTKTADAVQISGVQPVVFGWYAEHLPVRTPEPVRAEIGTDLVANAV	120
Qy	121	RAYVHFRSACVVDGCTALTFTTAVDGTGLIQGVIAIAPGLRTAVQSLHTGTAQLPLVPLAL	180
Db	121	RAYVHFRSACVVDGCTALTFTTAVDGTGLIQGVIAIAPGLRTAVQSLHTGTAQLPLVPLAL	180
Qy	181	PDSVLGKDTTHAVAGVWRTGLTFIVIRAMTAQCOKELGCRCAAVITGGLSLFSEVDPPP	240
Db	181	PDSVLGKDTTHAVAGVWRTGLTFIVIRAMTAQCOKELGCRCAAVITGGLSLFSEVDPPP	240

Db 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOQKELGCRCAAVITGGLSRLFSSEVDFFP 240

Qy 241 IDAQLTSLGLAHARLVPTSLPPATVSGSSGN 273

Db 241 IDAQLTSLGLAHARLVPTSLPPATVSGSSGN 273

RESULT 2

US-09-813-453A-7

Sequence 7, Application US/09813453A

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

TITLE OF INVENTION: ANTIBIOTICS

FILE REFERENCE: OG2-001

CURRENT APPLICATION NUMBER: US/09/813,453A

CURRENT FILING DATE: 2001-03-20

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 7

LENGTH: 255

TYPE: PRT

ORGANISM: Geobacter sulfurreducens

US-09-813-453A-7

Query Match 26.9%; Score 371.5; DB 9; Length 255;

Best Local Similarity 34.7%; Pred. No. 5e-27;

Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;

Qy 1 MLLIDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 2 LLDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Qy 61 RDAFISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 120

Db 58 RAVLISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 115

Qy 121 AAYVHFRSACVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLAL 180

Db 116 AGYEKRTSLIIVDFGTATTFTDYNRKGEYCGGAIAPGLVISTEALFORASKLPRVDIIR 175

Qy 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOQKELGCRCAAVITGGLSRLFSSE--VDFFP 239

Db 176 PSARIARTVMSQAGIYGVGLVDEIVTRKAEKSDAPRVATGGLASLAPESKTIE 235

Qy 240 PIDAQLTSLGL 250

Db 236 AVEEVLTLLEGL 246

RESULT 3

US-09-813-453A-55

Sequence 55, Application US/09813453A

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

TITLE OF INVENTION: ANTIBIOTICS

FILE REFERENCE: OG2-001

CURRENT APPLICATION NUMBER: US/09/813,453A

CURRENT FILING DATE: 2001-03-20

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 55

LENGTH: 256

TYPE: PRT

ORGANISM: Clostridium difficile

US-09-813-453A-55

Query Match 25.7%; Score 354; DB 9; Length 256;

Best Local Similarity 34.1%; Pred. No. 2.2e-25;

Matches 87; Conservative 55; Mismatches 99; Indels 14; Gaps 6;

Qy 1 MLLIDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 59

Db 2 LLDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 56

Qy 60 RDAFISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 119

Db 57 IRGIISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 114

Qy 120 VAAVHFRSACVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLA 179

Db 115 VAGIEKYGAPSLIVDFGTATTFTCAISEKGEYLGTTAPGKISSEALFQSAKSLPRVELA 174

Qy 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOQKELGCRCAAVI--TGGLSRLFSSE--- 235

Db 175 KPGMTICKSTVSAMQSGIYGVGLVDKIISIMKELNCDDVYKVIATGGLAKLIASETKS 234

Qy 236 VDFPPIDQLTSLGL 250

Db 235 IDY--VDGFLTLEGL 247

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 55

; LENGTH: 256

; TYPE: PRT

; ORGANISM: Clostridium difficile

US-09-813-453A-55

Query Match 25.7%; Score 354; DB 9; Length 256;

Best Local Similarity 34.1%; Pred. No. 2.2e-25;

Matches 87; Conservative 55; Mismatches 99; Indels 14; Gaps 6;

Qy 1 MLLIDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 59

Db 2 LLDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 56

Qy 60 RDAFISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 119

Db 57 IRGIISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 114

Qy 120 VAAVHFRSACVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLA 179

Db 115 VAGIEKYGAPSLIVDFGTATTFTCAISEKGEYLGTTAPGKISSEALFQSAKSLPRVELA 174

Qy 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOQKELGCRCAAVI--TGGLSRLFSSE--- 235

Db 175 KPGMTICKSTVSAMQSGIYGVGLVDKIISIMKELNCDDVYKVIATGGLAKLIASETKS 234

Qy 236 VDFPPIDQLTSLGL 250

Db 235 IDY--VDGFLTLEGL 247

RESULT 4

US-09-813-453A-49

Sequence 49, Application US/09813453A

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

TITLE OF INVENTION: ANTIBIOTICS

FILE REFERENCE: OG2-001

CURRENT APPLICATION NUMBER: US/09/813,453A

CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 49

LENGTH: 258

TYPE: PRT

ORGANISM: Bacillus stearothermophilus

US-09-813-453A-49

Query Match 24.8%; Score 342.5; DB 9; Length 258;

Best Local Similarity 31.0%; Pred. No. 2.7e-24;

Matches 78; Conservative 60; Mismatches 105; Indels 9; Gaps 4;

Qy 1 MLLIDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 59

Db 2 IFVLDVGNHSHVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 56

Qy 60 RDAFISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 119

Db 57 IRGIISSVVPVLTITADAVAQISGVQVVPFGPWAYEHLVPRIPPEVRAEIGTDLVANAV 114

Qy 120 VAAVHFRSACVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLA 179

Db 115 VAGIHLGSPLIIVDFGTATTFTYINERKQYMGGAIPGIMISTEALFARAALKPRIEIA 174

Qy 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOQKELGCRCAAVITGGLSRLFSSEVD-F 238











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 22.266 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVFGIQENG.....ARLVPTSLPPATVSGSSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL.21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1379	100.0	273	16	O83446
2	349.5	25.3	259	16	O8XHL5
3	348.5	25.3	255	16	O8R7M2
4	345.5	25.1	259	16	O92F54
5	344.5	25.0	259	16	O8YAC5
6	337.5	24.5	261	16	O9A621
7	329.5	23.9	258	2	O9F985
8	320.5	23.2	254	16	O9KGH5
9	316.5	23.0	273	16	O97EB4
10	291.5	21.1	265	16	O9X8N6
11	285	20.7	246	16	O9W2Y5
12	267	19.4	274	16	O9CD56
13	260.5	18.9	256	16	O8RFE4
14	254	18.4	272	16	O06282
15	248	18.0	262	16	O9RX54
16	192.5	14.0	212	2	O32514

17	170	12.3	262	16	O51477
18	151.5	11.0	224	16	O98Q93
19	147.5	10.7	592	16	O9JXF1
20	147.5	10.7	592	16	O9JW17
21	136.5	9.9	257	16	P74045
22	136	9.9	248	16	O9HWC1
23	132	9.6	276	16	O8YQD7
24	127.5	9.2	295	16	O8Y2M4
25	122.5	8.9	229	16	O67753
26	105	7.6	380	17	O9HRS8
27	105	7.6	430	16	O9A4D7
28	104.5	7.6	5532	2	O9S0R4
29	104	7.5	1456	2	O9E636
30	103.5	7.5	3972	2	O9S0R8
31	98	7.1	3192	2	O9L4W4
32	96.5	7.0	569	10	O9AXZ3
33	95.5	6.9	841	16	O9ACU8
34	95	6.9	366	2	O8VP07
35	95	6.9	613	2	O93LN4
36	94	6.8	383	16	O53382
37	93.5	6.8	469	16	O98K12
38	93	6.7	335	16	O9KEK9
39	93	6.7	347	16	O9RV45
40	93	6.7	638	2	O9L8K5
41	92	6.7	504	10	O24592
42	91.5	6.6	252	10	O94JD6
43	91.5	6.6	327	16	O98P49
44	91.5	6.6	2512	16	O10896
45	91.5	6.6	4077	2	O52820

ALIGNMENTS

RESULT 1  
O83446 PRELIMINARY; PRT; 273 AA.  
ID O83446  
AC O83446;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DE '01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein TP0431.  
GN TP0431.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOLS;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T., McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;  
RA "Complete genome sequence of Treponema pallidum, the syphilis agent, from a human patient."  
RT Spirochaete.  
RL Science 281:375-388(1998).  
DR EMBL; AF001220; AAC65417.1; -.  
DR TIGR; TP0431; -.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 100.0%; Score 1379; DB 16; Length 273;  
Best Local Similarity 100.0%; Pred No. 5.4e-98;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLLIDVGNHVFGIQENGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRSL 60





QY 1 MLLIDVGNHVVFGIENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
 DB 3 LLAIEQNTNTMFAT---HDGASWVAQ--WRSATESTRADAYVWVLSQGLGFRAI 58  
 QY 61 RDAFTSSVVPVLTKTADAVAOISGVQVPGVPGWAEHLFVRPEVRAEIGTDLVANAV 120  
 DB 59 DAVIISVVQSFNFNLSRRYFNVPLVGENAKLGDIVRIEKP--SEAGADRLVNAI 116  
 QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLAL 180  
 DB 117 GAAMVPGVLVIDSTATFTDIVADGAFEGGIAPGINSQALHEAAKLPRIAIOR 176  
 QY 181 P--DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRFSEYD- 237  
 DB 177 PAGNRIVGTDTSVMSQSGVFWGYSISLIEGLVARIKAEGERPMTVIATGGVASLFEATDS 236  
 QY 238 FPPIDAQLTSLGLAHAR 255  
 DB 237 IDHFDSDLTIRGLLEIYR 254

## RESULT 7

Q9F985

ID Q9F985 PRELIMINARY; PRT; 258 AA.

AC Q9F985; QUERY MATCH 23.9%; SCORE 329.5; DB 2; LENGTH 258;  
 DT 01-MAR-2001 (TREMELREL. 16, Created)  
 DT 01-MAR-2001 (TREMELREL. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DE Putative 32 kDa replication protein.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Geobacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V;  
 RA Vasquez C., Pichuanes S., Saavedra C.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF198621; AAG28531.1; --  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMS; TIGR00671; baf; 1.  
 SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 23.9%; Score 329.5; DB 2; Length 258;  
 Best Local Similarity 30.3%; Pred. No. 1.2e-17;  
 Matches 77; Conservative 60; Mismatches 104; Indels 13; Gaps 5;

DB 1 MLLIDVGNHVVFGI--QGENGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 59  
 DB 2 IFVLVDGNTNTVLGYDGDDE----LKHWRITSGRGTDEYGMTIKALLNHVGLQFSD 56  
 QY 60 LEDAFISSVVPVLTKTADAVAOISGVQVPGVPGWAEHLFVRPEVRAEIGTDLVANA 119  
 DB 57 IDGIITISSVVPVLPFALEMRCKLYFKPIIYVPGIKTGLNLIKNDNP--REVGAADRVNA 114  
 QY 120 VAAVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLA 179  
 DB 115 VAGIHLGSPLIIVDFGATTTCYINEHKQYMGGAIPGIMISTEALFAAKLPRIEIA 174  
 QY 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRFSE--V 236  
 DB 175 RPDDIIGNTVSAMQAGILYGVGVGVEGIVSRMKAKSPVPKVIATGGLASLASESNI 234  
 QY 237 FPPIDAQLTSLGL 250  
 DB 235 DI--VDPFLTLTGL 246

## RESULT 8

Q9KGHS

ID Q9KGHS PRELIMINARY; PRT; 254 AA.

AC Q9KGHS; QUERY MATCH 23.2%; SCORE 320.5; DB 16; LENGTH 254;  
 DT 01-OCT-2000 (TREMELREL. 15, Created)  
 DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DE Hypothetical protein BH0086.  
 GN BH0086.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RL halodurans and genomic sequence comparison with Bacillus subtilis.";  
 DR EMBL; AP01507; BAB03805.1; --  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMS; TIGR00671; baf; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 23.2%; Score 320.5; DB 16; Length 254;  
 Best Local Similarity 31.4%; Pred. No. 5.7e-17;  
 Matches 80; Conservative 57; Mismatches 103; Indels 15; Gaps 5;  
 QY 1 MLLIDVGNHVVFGIENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
 DB 2 ILVIDVGNTNTVLGYDQDE---TLVHWRLATSRKTEDEYAMTVRSFLDHAGLQFQDI 57  
 QY 61 RDAFTSSVVPVLTKTADAVAOISGVQVPGVPGWAEHLFVRPEVRAEIGTDLVANAV 120  
 DB 58 DGIVTSSVVPVPMFSLQMKCKYFHTVPMIIGFGIKTGLNLIKNDNP--KEVGADRIYNAV 115  
 QY 121 AAVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTQALPLVPLAL 180  
 DB 116 AAIELYGYPAIVVDFGTATYCLINEKKQYAGGVAGGIMISTEALYHRASKLPRIEIAK 175  
 QY 181 PDSVLGKDTTHAVQAGVVRGTLF---VIRAMIAQCQKELGCRCAAVITGGLSRFSEY 236  
 DB 176 PKQVGTNTIDSMQSGIFVGYVQVGVWVRMKQAQAESE----PKVIATGGLAKLICTES 231  
 QY 237 D-FPPIDAQLTSLGL 250  
 DB 232 ETIDVIDSFTLTKGL 246

## RESULT 9

Q97EB4

ID Q97EB4 PRELIMINARY; PRT; 273 AA.

AC Q97EB4; QUERY MATCH 23.2%; SCORE 320.5; DB 16; LENGTH 254;  
 DT 01-OCT-2001 (TREMELREL. 18, Created)  
 DT 01-OCT-2001 (TREMELREL. 18, Last sequence update)  
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DE Predicted transcriptional regulator, homolog of Bvg accessory  
 DE factor.  
 GN CAC3200.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,



```

RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007815; AAK81136.1; -
DR InterPro; IPR004619; Baf.
DR InterPro; IPR000515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 30331 MW; DE7B4D5923E72842 CRC64; *

Query Match 23.0%; Score 316.5; DB 16; Length 273;
Best Local Similarity 30.6%; Pred. No. 1.2e-16;
Matches 78; Conservative 59; Mismatches 107; Indels 11; Gaps 4;

QY 1 MLLIDVGNHVVFGIQTGNGRCVRELFRLAPDARKTQDEYSLIIHALCERAGVGRASL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 ILVLDVGNTHVILGIYNDTK----LTAEWRLSTDLRSADIEGIVQMFLFQDDKLDPLV 68
61 RDAFISVVVPLTITADAVAQISGVQPVVFGPWAYEHLVPRIPEVRAEIGTDLVANAV 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 EGVIISSVVPNIMYSLEHMIRKYEKINPLVVGPGIKTGINKYDNP--KEVGADRIVANV 126
121 AAYVHRSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTALPLVPLAL 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 AAHEYKRLSLIIDFTATTTCVAVRENGDYLGGATCPGKIVSSSEALFEKAALPRVELIK 186
181 PDSVLGKDTTHAVQGVVGRFL---FVIRAMIAQCOKELCRAAVITGLSLRFSSEV 236
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 PAYAIKNTIISGIVGIVGVYIYVERKKEQLQEKEPLVAVTGGLAKLTISEA 246
237 -DFPPIDAQLTSLGL 250
:|||||
247 KNVDVINPFLTLEGL 261

RESULT 10
Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
SS Streptomyces coelicolor.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX NCBI_TaxID=1902;
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 mb streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;

```

```

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL049628; CAB40880.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein.
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 21.1%; Score 291.5; DB 16; Length 265;
Best Local Similarity 33.1%; Pred. No. 9.9e-15;
Matches 89; Conservative 47; Mismatches 108; Indels 25; Gaps 9;

QY 1 MLLIDVGNHVVFGI-QGNGRCVRELFRLAPDARKTQDEYSLIIHALCERAGVGRAS 59
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 LITIDVGNTHVILGLFDGED-----IVEHWRISTDSRTADELAVLLQGLMGHPLIGDE 56
60 LRDAF-----ISSVWPVLTKTIADAVAQISG-VQPVVFGPWAYEHLVPRIPEVRAEIGT 114
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 LCDGIDGTAICATVSVLHELREVTTRYGDVPAVLVEPGVKTGVILTDPH--KEVGAD 114
115 LVANAVAAVHFRSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTALP 174
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 RIINAVAAVELYGGPAIVDFGTATTDAVSARGEYIGGVIAPIGIEISVEALGVKGAQLR 174
175 LVPLALPDSVLGKDTTHAVQGVVGRFTLVIRAMIAQCOKELG--CRCAAVITGLSLRL 231
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 KIEVARPSRVIGKNTVEAMQSGIVYGFAGQGVGNRMARELADDDPDDVTVIATGLAPM 234
232 F---SSEVDFPPIIDAQLTSLGLAHIALRV 257
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 VLGESSVID--EHEPWLITLMLG---RLV 257

RESULT 11
Q9WZY5 PRELIMINARY; PRT; 246 AA.
ID Q9WZY5
AC Q9WZY5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
GN TM0883.
OS Thermotoga maritima.
OX Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.E., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL: AE001754; AAD35964.1; -
DR TIGR; TM0883; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.97518 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFGIQENG.....ARLVPTSLLPATVSGSGGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 262574 seqs, 29422922 residues

total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92.5	6.7	351	4	US-09-632-947B-11
2	91	6.6	829	4	US-09-413-814-105
3	87.5	6.3	3519	4	US-09-428-517-4
4	83.5	6.1	1049	3	US-08-772-270A-11
5	83.5	6.1	1244	5	PCT-US93-10500-2
6	82	5.9	922	4	US-08-460-269C-6
7	82	5.9	4150	4	US-09-428-517-2
8	81.5	5.9	437	2	US-08-989-925-3
9	81.5	5.9	2532	4	US-09-215-694-10
10	81	5.9	826	4	US-09-564-803-224
11	81	5.9	826	4	US-09-564-803-226
12	80.5	5.8	437	1	US-08-764-343-1
13	80.5	5.8	683	4	US-09-134-001C-5576
14	80.5	5.8	819	4	US-09-651-658-15
15	80.5	5.8	819	4	US-09-650-855-15
16	80.5	5.8	5087	4	US-09-144-085-1
17	80	5.8	434	2	US-08-989-925-1
18	79	5.7	712	4	US-09-708-426-9
19	78.5	5.7	801	1	US-08-725-012-2
20	78.5	5.7	826	4	US-09-564-805-2
21	78	5.7	869	4	US-09-071-035-264
22	78	5.7	1638	4	US-09-071-035-258
23	78	5.7	1638	4	US-09-071-035-262
24	78	5.7	1638	4	US-09-071-035-266
25	76.5	5.5	418	4	US-09-030-267-5
26	76.5	5.5	3816	4	US-09-428-517-3
27	76	5.5	161	2	US-08-403-852D-22

28	76	5.5	161	3	US-08-510-646B-23	Sequence 23, Appl
29	76	5.5	161	4	US-09-231-818-22	Sequence 22, Appl
30	75.5	5.5	320	4	US-09-036-987A-14	Sequence 14, Appl
31	75.5	5.5	320	4	US-09-370-700-14	Sequence 14, Appl
32	75.5	5.5	1841	2	US-08-804-227C-6	Sequence 6, Appl
33	75.5	5.5	4545	2	US-08-804-227C-14	Sequence 14, Appl
34	75.5	5.5	4550	2	US-08-804-227C-8	Sequence 8, Appl
35	75.5	5.5	4550	2	US-08-804-198-2	Sequence 2, Appl
36	75	5.4	2595	4	US-09-036-987A-2	Sequence 2, Appl
37	75	5.4	2595	4	US-09-370-700-2	Sequence 2, Appl
38	75	5.4	5215	4	US-09-105-537-2	Sequence 2, Appl
39	75	5.4	5588	4	US-09-036-987A-6	Sequence 6, Appl
40	75	5.4	5588	4	US-09-370-700-6	Sequence 6, Appl
41	74.5	5.4	1562	3	US-09-320-878-3	Sequence 3, Appl
42	74.5	5.4	1562	4	US-09-105-537-35	Sequence 35, Appl
43	74.5	5.4	11877	4	US-09-105-537-6	Sequence 6, Appl
44	74	5.4	565	4	US-08-952-365-8	Sequence 8, Appl
45	73.5	5.3	1580	2	US-08-804-227C-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-632-947B-11  
; Sequence 11, Application US/09632947B  
; Patent No. 6356845  
; GENERAL INFORMATION:  
; APPLICANT: Pharmacia & Upjohn  
; TITLE OF INVENTION: CRYSTALLIZATION AND STRUCTURE DETERMINATION OF  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS  
; TITLE OF INVENTION: UDP-N-ACETYLENOLPYRUVYLGLUCOSAMINE REDUCTASE (S. aureus  
; TITLE OF INVENTION: MuB)  
; FILE REFERENCE: 268,6241 0101  
; CURRENT APPLICATION NUMBER: US/09/632,947B  
; CURRENT FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/147,164  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Bordetella pertussis  
; US-09-632-947B-11

Query Match	6.7%	Score 92.5	DB 4	Length 351
Best Local Similarity	24.0%	Pred. No. 0.041		
Matches	68	Conservative 36	Mismatches 94	Indels 85
Gaps	15			
QY	31	LAPDARKTQDEYSLLIHALCERAGVGRASLRDAFIS----	SVVPVLTKTIADAVAQISGV	86
Db	11	VAPLAPQAQD-----LRCFNT-LGLASHAPAFVALTEPSQLPALS-ALAPRFRQL---	58	
QY	87	QPVVFGWYAEHLPL-----VRPEPVRAEIGTDLVANAAVY-----	HFR	127
Db	59	--VVLGGGNNVLPASIDGLVAQVRLP-----GVRLVGCADAWYVAAAGENWHGFV	109	
QY	128	SACVVVDCGTALTFTTAVDGTGLGQVAIPGLRTAVQSLHTGTAGLPVLPALPDSVLGK	187	
Db	110	TAC--VDNGW-----DG---LENLALPGTVGAAPVQNGAYGVELA-----	146	
QY	188	DTTHAVQAGVVRGTLEFIRAMIAQCO-----KELGCRCAAVITGGLSRLFSSEVD	237	
Db	147	DRFHSLTANDVKGGRWY-EMGAECRFAYEDSFEKHQEPGAWVIGSVYRFALPRWQPVLD	205	
QY	238	FPPIDQQLTSLGLAHIAHRLVPTSL-----LPPATVSGSGN	273	
Db	206	YPDQORHAALDGAAPTARAVYDAVCAIRRAKLPDPAPVGNAGS	248	

RESULT 2  
US-09-413-814-105



;; INFORMATION FOR SEQ ID NO: 11:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1049 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Actinobacillus pleuropneumoniae  
;; US-08-772-270A-11

Query Match 6.1%; Score 83.5; DB 3; Length 1049;  
Best Local Similarity 20.9%; Pred. No. 2.2; Indels 41; Gaps 9;  
Matches 53; Conservative 47; Mismatches 112;  
QY 26 RELFLAPDARKTQDEYSLIIHALCERAGVGRASLRDAFI---SSVVPVLTITADAVAQ 82  
Db 93 RNELEVAHKSGLTADQFL---GLTERGLTLFAPQLDQFLQKHSKISNVVGSSTGDAVSK 148  
Y 83 ISGVQPVVPGPWAYEHLVPRIPVRAEIGTDLVA-NAVAAYVHFRSACVVVDCGTALTF 141  
Db 149 LAKSQTIIISG-----IQSVLGTVLGINLNEAIISGGSELEAEAGVSLAS 194  
QY 142 TAVDGTGLIQGVAIAPGLRTAVQSL-----HTGTAQLPLVPLALPDSVLGKDTT 190  
Db 195 ELV--SNIAGTTTIDAFITQIFGKLAENAKGLGGVGRQLQNISGSALSKTGLGLDII 252  
QY 191 HAVQAGVVRGTLFVIRAMIAOCOKEL--GCRCAAVITGGLSRFSSEVDFFPPIIDAQTLT 248  
Db 253 SSLSGVTRS--FALRNKNASTSTKVAAGFELSNOVIGGITKAVSSVYLAQRLRAGLSTT 310  
QY 249 GLAHIALRVPTSL 261  
Db 311 GPA--AALIASSI 321

RESULT 5  
PCT-US93-10500-2

;; Sequence 2, Application PC/TUS9310500  
;; GENERAL INFORMATION:  
;; APPLICANT: Chang, Yung-Fu  
;; TITLE OF INVENTION: Recombinant Vaccine For Procine  
;; TITLE OF INVENTION: Pleuropneumoniae  
;; NUMBER OF SEQUENCES: 2  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Alan S. Korman  
;; STREET: 1600 Empire Tower  
;; CITY: Buffalo  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 14202  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/10500  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/972,229  
;; FILING DATE: 05-NOV-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Korman, Alan S.  
;; REGISTRATION NUMBER: 33,932  
;; REFERENCE/DOCKET NUMBER: 19603/00001  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 716-853-8104  
;; TELEFAX: 716-853-8109  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1244 amino acids

;; TYPE: amino acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; FRAGMENT TYPE: N-terminal  
;; ORIGINAL SOURCE:  
;; ORGANISM: Actinobacillus pleuropneumoniae  
;; STRAIN: Serotypes 2, 3, 4, 6 and 8  
;; INDIVIDUAL ISOLATE: Swine  
;; CELL TYPE: Gram negative bacterium  
;; PCT-US93-10500-2

Query Match 6.1%; Score 83.5; DB 5; Length 1244;  
Best Local Similarity 20.9%; Pred. No. 2.9;  
Matches 53; Conservative 47; Mismatches 112; Indels 41; Gaps 9;  
QY 26 RELFLAPDARKTQDEYSLIIHALCERAGVGRASLRDAFI---SSVVPVLTITADAVAQ 82  
Db 266 RNELEVAHKSGLTADQFL---GLTERGLTLFAPQLDQFLQKHSKISNVVGSSTGDAVSK 321  
QY 83 ISGVQPVVPGPWAYEHLVPRIPVRAEIGTDLVA-NAVAAYVHFRSACVVVDCGTALTF 141  
Db 322 LAKSQTIIISG-----IQSVLGTVLGINLNEAIISGGSELEAEAGVSLAS 367  
QY 142 TAVDGTGLIQGVAIAPGLRTAVQSL-----HTGTAQLPLVPLALPDSVLGKDTT 190  
Db 368 ELV--SNIAGTTTIDAFITQIFGKLAENAKGLGGVGRQLQNISGSALSKTGLGLDII 425  
QY 191 HAVQAGVVRGTLFVIRAMIAOCOKEL--GCRCAAVITGGLSRFSSEVDFFPPIIDAQTLT 248  
Db 426 SSLSGVTRS--FALRNKNASTSTKVAAGFELSNOVIGGITKAVSSVYLAQRLRAGLSTT 483  
QY 249 GLAHIALRVPTSL 261  
Db 484 GPA--AALIASSI 494

RESULT 6

US-08-460-269C-6  
;; Sequence 6, Application US/08460269C  
;; Patent No. 6197548  
;; GENERAL INFORMATION:  
;; APPLICANT: CLARE, JEFFREY J.  
;; ROMANOS, MICHAEL A.  
;; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST  
;; NUMBER OF SEQUENCES: 17  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.  
;; STREET: 2200 Clarendon Blvd., Suite 1400  
;; CITY: ARLINGTON  
;; STATE: VA  
;; COUNTRY: USA  
;; ZIP: 22201  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/460,269C  
;; FILING DATE: 02-Jun-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Lebovitz, Richard M.  
;; REGISTRATION NUMBER: 37,067  
;; REFERENCE/DOCKET NUMBER: Popov-2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 243-6333  
;; TELEFAX: (703) 243-6410  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:

LENGTH: 922 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-08-460-269C-6

Query Match 5.9%; Score 82; DB 4; Length 922;  
Best Local Similarity 25.2%; Pred. NO. 2.7;  
Matches 54; Conservative 18; Mismatches 90; Indels 52; Gaps 10;

QY 78 DAVAQISGVQPVVFGPWAYEHLFVRIPEPVRAEIGTDLVANAAVHFRSACVVVDCGT 137  
DB 104 EGVRELGIVTVYKAGLVADHATIANVSDTRDDG-----IALIVAGEQA-----QA 150  
QY 138 ALTFRAVDTGLIQGVAIAPGLRTAVQ-----SLHTGTALQPLVPLALPDS--VLGKD 188  
DB 151 STADSTLQAG--GVRVERGANVTVORSTIVDGLHIGTLQ-PLQPEDLPPSRVVLGDT 206  
QY 189 TTHAVQAVVVGTLFVIRAMIAQCQKEL-----GRCQA--AVITGGLSRLESSEVDF 238  
DB 207 SVTAVPASGAPAAVFEVGA-----NELTVDGCHITGGAAGVAAAMDGAIVHLQRATI-- 258  
QY 239 PPIDAQLTSLGLAHIAHRLVPTSLPPATVYSGSG 272  
DB 259 -----RRGAPAGGAVPGGAVPGGAVPGGFG 284

RESULT 7  
US-09-428-517-2  
Sequence 2, Application US/09428517  
Patent No. 6251636  
GENERAL INFORMATION:  
APPLICANT: Betlach, Mary C.  
APPLICANT: Shah, Sanjay Krishnakant  
APPLICANT: McDaniel, Robert  
APPLICANT: Tang, Li  
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 30062-20029.00  
CURRENT APPLICATION NUMBER: US/09/428,517  
CURRENT FILING DATE: 1999-10-28  
EARLIER APPLICATION NUMBER: 60/120,254  
EARLIER FILING DATE: 1999-02-16  
EARLIER FILING DATE: 1998-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4150  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
OTHER INFORMATION: Oleandolide PKS  
US-09-428-517-2

Query Match 5.9%; Score 82; DB 4; Length 4150;  
Best Local Similarity 21.7%; Pred. NO. 26;  
Matches 68; Conservative 31; Mismatches 94; Indels 120; Gaps 15;  
QY 13 FGIQENGRCVRCVRELFRAPDTPARKTQDYISLLIHLCERAGVGRASLRDAFISVVVP-- 70  
DB 1463 FGISGTN-AHVIVEE-----APAEPEPEPE-----RGPGS-----VGVVVPVW 1499  
QY 71 -----VLTKTADIADVAQISGVQPVVFGPWAA-----YEHLPVRIPEPVRAEIGTDLV 116  
DB 1500 VSGRDAGALREQAARLAHVSQVAVDVG--WSLNVTRSVFHEHRAVNVGSEL-----DAM 1552  
QY 117 ANAAAYVHFRSACVVVDCGTALTETAVDGTGLIQGVAIAPGLRTAV-----QSLHNG 169  
DB 1553 AESLAG---FRAGGV-----PGVVGVAPEGRVVRVFPFGGSGQVWNA 1595  
QY 170 TQALFLVPL-----ALPDSVLGKDTTHAVQAG-----VVRGTLFVIRAMIAQ 211

DB 1596 AGLLDACPVPFAEVAECAVLDPLTGWSLVEVLRGGGEAVLGRVDVQPALWAVVSLAR 1655  
QY 212 CQKELGRCRAV-----ITGGLSRLESSEVDFPPIDAQLTSLGLAHIAHRLVPT 259  
DB 1656 TWRYGVPEPAVVGHSQGEITAAACVAGLS-----LADGARVVVL 1695  
QY 260 SLLPPATVYSGSG 272  
DB 1696 RSRARIAAGGG 1708

RESULT 8  
US-08-989-925-3  
Sequence 3, Application US/08989925  
Patent No. 5989820  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA: 2.0  
APPLICATION NUMBER: US/08/989,925  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0440 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1806040  
US-08-989-925-3

Query Match 5.9%; Score 81.5; DB 2; Length 437;  
Best Local Similarity 23.5%; Pred. NO. 1;  
Matches 48; Conservative 28; Mismatches 69; Indels 59; Gaps 10;  
QY 38 TQDEYSLTHALCERAGVGRASLRDAFISVWPVLTKTIADVAQISGVQBP--VVFQW 95  
DB 36 TKDQPYL-KSCVCEAENGKTTITSVAMTSALPIOK-----LEPQIAVANTYA 83  
QY 96 YEHLPVRIPE--PVRAEIGTDLVANAAVHFRSACVVVDCGTALTETAVDGTGLI 153  
DB 84 CKGLD-RIBERFILNQPSTQIVANAKA-----VTGAKDAVTTT----- 124



QY 154 AIAPLGRTAVQSLHTGTAQLPLVPLALPDSVLGKDTTHAVQAVGVRGTLFVIRAMTAQOQ 213  
Db 125 -----GAKSVASTIGV-----MDKTKGAVTGSVEKTSVVGSGSI---N 161  
QY 214 KELGRCRAAVITGGL-SRLFSSEV 236  
Db 162 TVLGSRMQLVSSGVENALTKSEL 185

RESULT 9

US-09-215-694-10  
; Sequence 10, Application US/09215694B  
; Patent No. 6391583  
; GENERAL INFORMATION:  
; APPLICANT: Wisconsin Alumni Research Foundation  
; APPLICANT: Hutchinson, Charles R.  
; APPLICANT: Kennedy, Jonathan n.m.i  
; APPLICANT: Park, Cheonseek n.m.i  
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS  
; FILE REFERENCE: 960296.95718  
; CURRENT APPLICATION NUMBER: US/09/215,694B  
; CURRENT FILING DATE: 1999-12-18  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 2532  
; TYPE: PRT  
; ORGANISM: Aspergillus terreus  
US-09-215-694-10

Query Match 5.9%; Score 81.5; DB 4; Length 2532;  
Best Local Similarity 21.1%; Pred. No. 14;  
Matches 38; Conservative 29; Mismatches 68; Indels 45; Gaps 7;  
QY 6 VGNHVVFGIOGNGRVCVRLFRAPDARKTDQ-EYSLIHALCERAGVGRASLRDAF 64  
Db 64 VGSNHRVGHFLDEDPALFADSFNNSTEVASCMDPQYLLIEVYE-----AL 112  
QY 65 ISSVVPVLTKTADAVQISGVQVVPVFGWAY-----EHLVPVRIPEP-VRAEIGTDLV 116  
Db 113 EAAGIP-----LEQVSGSKTGVFAGTMYHDYQGSFQRPQEPALPRYFITGNAGT-ML 162  
QY 117 ANAVAAVHFRSACVVDGCTALTFTAVDGTGLIOGVAIPGLRTAVQSLHTGTAQLPLV 176  
Db 163 ANRVSHFYDLRGPVSIDTACSTTLTA-----LHLAIOSLRAGESDMAIV 207

RESULT 10

JS-09-564-805-224  
; Sequence 224, Application US/09564805  
; Patent No. 6333403  
; GENERAL INFORMATION:  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.F.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Rommens, Johanna M.  
; APPLICANT: Myriad Genetics, Inc.  
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility  
; TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes  
; FILE REFERENCE: 2318-258  
; CURRENT APPLICATION NUMBER: US/09/564,805  
; CURRENT FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: US 60/107,468  
; PRIOR FILING DATE: 1998-11-06  
; PRIOR APPLICATION NUMBER: 09/434,382  
; PRIOR FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 224  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Pan troglodytes  
US-09-564-805-224

Query Match 5.9%; Score 81; DB 4; Length 826;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 47; Conservative 19; Mismatches 68; Indels 68; Gaps 8;  
QY 26 RELFRLAP-----DARKTODEYSLIHALCERAGVGRASLRDAFTSSV----- 68  
Db 202 RPLSLSPERSSDSENE-PLPHGVQRRGRVDRSSLVVAVFICKLHKKRGNFLVLKAK 260  
QY 69 ---VPVLTKTADAVQISGVQVVPVFGWAYEHLVPVRIPEPVRARIEIGTDLVANAAVAVH 125  
Db 261 EMGLPVGTAATAPIIAAVKDKSI-----THEGREILAEEL----- 296  
QY 126 FRSAVVVDGCTALTFTAVDGTGLIOGVAIPGLRTAVQSLHTGTAQLP--LVPLALPDS 183  
Db 297 ----CTPPDPGAFFVVECPDESFIQ-----PICENATFORQOGKADAPVALVYHMAPES 347  
QY 184 VL-----GKDTTHAV 193  
Db 348 VLVDSTRYQQWMERFGPDTHLV 369

RESULT 11

US-09-564-805-226  
; Sequence 226, Application US/09564805  
; Patent No. 6333403  
; GENERAL INFORMATION:  
; APPLICANT: Tavtigian, Sean V.  
; APPLICANT: Teng, David H.F.  
; APPLICANT: Simard, Jacques  
; APPLICANT: Rommens, Johanna M.  
; APPLICANT: Myriad Genetics, Inc.  
; TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility  
; TITLE OF INVENTION: Gene and a Paralog and Orthologous Genes  
; FILE REFERENCE: 2318-258  
; CURRENT APPLICATION NUMBER: US/09/564,805  
; CURRENT FILING DATE: 2000-05-05  
; PRIOR APPLICATION NUMBER: US 60/107,468  
; PRIOR FILING DATE: 1998-11-06  
; PRIOR APPLICATION NUMBER: 09/434,382  
; PRIOR FILING DATE: 1999-11-05  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 226  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Gorilla gorilla  
US-09-564-805-226

Query Match 5.9%; Score 81; DB 4; Length 826;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 47; Conservative 19; Mismatches 68; Indels 68; Gaps 8;  
QY 26 RELFRLAP-----DARKTODEYSLIHALCERAGVGRASLRDAFTSSV----- 68  
Db 202 RPLSLSPERSSDSENE-PLPHGVQRRGRVDRSSLVVAVFICKLHKKRGNFLVLKAK 260  
QY 69 ---VPVLTKTADAVQISGVQVVPVFGWAYEHLVPVRIPEPVRARIEIGTDLVANAAVAVH 125  
Db 261 EMGLPVGTAATAPIIAAVKDKSI-----THEGREILAEEL----- 296  
QY 126 FRSAVVVDGCTALTFTAVDGTGLIOGVAIPGLRTAVQSLHTGTAQLP--LVPLALPDS 183  
Db 297 ----CTPPDPGAFFVVECPDESFIQ-----PICENATFORQOGKADAPVALVYHMAPES 347  
QY 184 VL-----GKDTTHAV 193  
Db 348 VLVDSTRYQQWMERFGPDTHLV 369

RESULT 12

US-08-764-343-1  
; Sequence 1, Application US/08764343

```

213 QXELGCRCAAVITGGGL-SRLFSEV 236
      || | : : : | : ||:
161 NTVLGRMMQLVSSGVENALTKSEL 185

QY
Db

RESULT 13
US-09-134-001C-5576
; Sequence 5576, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.74823 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453a-10

Perfect score: 1379

Sequence: 1 MLLIDVGNHVFIQIENG.....ARLVPTSLPPATVSGSGSN 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	288	20.9	233	1	YACB_BACSU
2	106.5	7.7	344	1	GUN4_TRIPE
3	98.5	7.1	679	1	MYO1_YEAST
4	93.5	6.8	335	1	MEH1_BACSU
5	93.5	6.8	459	1	COBG_PSEDE
6	92.5	6.7	351	1	MURB_BORPE
7	91.5	6.6	497	1	GATA_MYCLE
8	90.5	6.6	290	1	MUR1_DEIRA
9	90.5	6.6	327	1	FMT_RALSO
10	90	6.5	747	1	GUND_CELFI
11	88.5	6.4	836	1	VG26_BPMD2
12	88	6.4	384	1	POQE_METEX
13	87.5	6.3	3519	1	OL56_STRAT
14	86	6.2	285	1	DPH2_HUMAN
15	85.5	6.2	489	1	CL28_MYCTU
16	85.5	6.2	545	1	DCIP_AZOBR
17	85	6.2	280	1	DHP1_MYCTU
18	85	6.2	351	1	ID12_RHILO
19	84.5	6.1	397	1	DP3B_MYCSM
20	84	6.1	270	1	ISPD_STRCO
21	84	6.1	562	1	ATKA_ALIAC
22	83.5	6.1	817	1	MUTS_THECA
23	83.5	6.1	1008	1	SN14_YEAST
24	83.5	6.1	1049	1	RP31_ACTPL
25	82.5	6.0	232	1	VG65_BPMD2
26	82.5	6.0	238	1	HIS4_METVA
27	82.5	6.0	327	1	GLK_2YMMO
28	82.5	6.0	338	1	GPDA_STRPY
29	82.5	6.0	381	1	MRP_MYCTU
30	82.5	6.0	468	1	HBM1_MYCTU
31	82.5	6.0	665	1	CVAL_RHIME
32	82	5.9	166	1	RS5_BACSU
33	82	5.9	302	1	COBV_PSEDE

34 82 5.9 415 1 DP42\_RHILO  
35 82 5.9 836 1 VG26\_BPML5  
36 82 5.9 922 1 PERT\_BORPA  
37 82 5.9 3354 1 CADN\_MOUSE  
38 81.5 5.9 277 1 MUR1\_BRUME  
39 81.5 5.9 437 1 ADPF\_HUMAN  
40 81.5 5.9 505 1 YLZ1\_CAEEL  
41 81.5 5.9 519 1 LEU1\_XYLFA  
42 81.5 5.9 633 1 DNAK\_AGR75  
43 81.5 5.9 1205 1 SMC\_MYCTU  
44 81 5.9 430 1 VG43\_BPPE3  
45 81 5.9 530 1 MUR1\_MYCLE

## ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:11-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita X., Lapido A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noack M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porvollik S., Prescott A.M.,  
RA Presecan E., Fujic P., Furnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche H., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
RA Sorokin M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Takeuchi M., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Tosato V., Wambuit R., Wedler E., Wedler H., Weitzenegger T.,  
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; D36185; BAA05305.1; -;  
 DR EMBL; Z99104; CAB11846.1; -;  
 DR Subtilisin; BG10133; YacB.  
 DR InterPro: IPR004619; Baf.  
 DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMS; TIGR00671; baf; 1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;  
 SQ  
 Query Match 20.9%; Score 288; DB 1; Length 233;  
 Best Local Similarity 29.7%; Pred. No. 8.6e-17;  
 Matches 62; Conservative 56; Mismatches 85; Indels 6; Gaps 3;  
 QY 1 MLLIDVGNHGVFGIQQENGRCVRELFRLAPDARKTODEYSLLIHALCERAGVGRASL 60  
 DB 2 LVIDVGNNTVLGV--HDGK--LEYHRIETSRHKTDEFGMLIRSLFDHSGLMFEQI 57  
 QY 61 RDAFISSVVPVLTKTITADVAQISGVQPVVFGPWAYEHLFVPEVRAEIGTDLVANAV 120  
 DB 58 DGIILSSVVPVIMFALERMCTKYFHEIPQVPGMKTGLNICYDNP--KEVGADRVNAV 115  
 QY 121 AAYVIFRSACVVDCGTALTFTAVDGTGLIQGVAIAPGLTAVQSLHTGTALPLPLAL 180  
 DB 116 AAHLHGNLVLVDFGTATYCYIDENKQYMGCAIAPGITISTEALYSRAAKLPRIETR 175  
 QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMI 209  
 DB 176 PNIIGKNVVSAMQSGILFYGQVEGIV 204  
 RESULT 2  
 ID GUN4\_TRIRE STANDARD; PRT; 344 AA.  
 AC O14405;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Endoglucanase IV precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase IV)  
 DE (Cellulase IV) (EGIV).  
 DE EGL4.  
 Trichoderma reesei (Hypocrea jecorina).  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreales; Hypocreaceae; Hypocrea.  
 NCBI\_TaxID=51453;  
 [1]  
 SEQUENCE FROM N.A., ACTIVITY, AND INDUCTION.  
 RC STRAIN-QM9414 / Rut C-30;  
 RX MEDLINE=98036137; PubMed=9370370;  
 RA Salohelmo M., Nakari-Setäläe T., Tenkanen M., Penttiläe M.;  
 RT "cDNA cloning of a Trichoderma reesei cellulase and demonstration of  
 RT endoglucanase activity by expression in yeast."  
 RL Eur. J. Biochem. 249:584-591(1997).  
 CC -!- FUNCTION: The biological conversion of cellulose to glucose  
 CC generally requires three types of hydrolytic enzymes:  
 CC (1) Endoglucanases which cut internal beta-1,4-glucosidic bonds;  
 CC (2) Exocellulohydrolases that cut the disaccharide cellobiose  
 CC from the nonreducing end of the cellulose polymer chain;  
 CC (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other  
 CC short cello-oligosaccharides to glucose.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- INDUCTION: By cellulose, cellobiose, lactose and sophorose.  
 CC -!- PTM: May also be O-glycosylated.

-----  
 CC -!- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.  
 CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN  
 CC (CBD).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; Y11113; CAA71999.1; -;  
 DR HSP; P00725; ZCBH.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR005103; Glyco\_hydro\_61.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR Pfam; PF03443; Glyco\_hydro\_61; 1.  
 DR ProDom; PD001821; CBD\_fungal; 1.  
 DR SMART; SM00236; fCBD; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 21 POTENTIAL  
 FT CHAIN 22 344 ENDOGLUCANASE IV.  
 FT DOMAIN 22 256 CATALYTIC (POTENTIAL).  
 FT DOMAIN 257 307 LINKER (POTENTIAL).  
 FT DOMAIN 308 344 CELLULOSE-BINDING (POTENTIAL).  
 FT DISULFID 315 332 BY SIMILARITY.  
 FT DISULFID 326 342 BY SIMILARITY.  
 FT CARBOHYD 80 80 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 344 AA; 35510 MW; 7FBFIC4AB705350C CRC64;  
 Query Match 7.7%; Score 106.5; DB 1; Length 344;  
 Best Local Similarity 23.2%; Pred. No. 0.11;  
 Matches 69; Conservative 28; Mismatches 89; Indels 111; Gaps 17;  
 QY 31 LAPDARKTODEYSLLIHALCERAGVGRASLR--DAFISSVVPVLTKTITADVAQISGVQ 88  
 DB 65 VSPDAQONPD---IICHKNATNA-KGHASVKGAGDTILFQWVPV----- 103  
 QY 89 VVEGPNWAYEHLFVRIPEVRAEIGTDLVANAVAYVHFRSACVVVDCGTALTFTAVDGTG 148  
 DB 104 ---PWPH-----FGPI-----VDYLANC-----NGDCEFTVD-KTLEFFKIDGVG 139  
 QY 149 LIQG-----VAIAPG---LRTAVQSLHT-----GTAQLP-LV 176  
 DB 140 LLSGGDPGTWASDVLSINNNNTWVVKIPDNLAPGNVYLRHLLIHLHSAQANGAQNYPQCF 199  
 QY 177 PLAL-----PDSVLGKDTTHAVQAGV-----VRGTLFV--IRAMIAQCKE 215  
 DB 200 NIAVSGSGSLQPSGLVGLTDLYHATDGLVNIYTSPLNIIPGPTVVSGLPTSVAQSSA 259  
 QY 216 LGCRCAAVITGGLSRFLSFSEVDPDPAQLTSLGLAHIALRVLTSLPPATVSSGS 272  
 DB 260 ATATASATVPGGSGGPTS-----RTTTTARTQAASSRPSS-TTPATTSPAG 305  
 RESULT 3  
 ID MTOL\_YEAST STANDARD; PRT; 679 AA.  
 AC P53070;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mitochondrial translation optimization protein.  
 GN MTOL OR IPSI OR YGL236C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NCBI\_TaxID=4932;  
 [1]  
 RP SEQUENCE FROM N.A.



QY 121 AAYVHRSACVVDGCTALTFTAYDGTGLIOGVA-----I 155  
 Db 153 -----VVVDIGGGTTEVALISFG---GVVSCHSIRIGGDOLDIEDIVSFVRKKYNL 199  
 QY 156 APGLRTAVQ-SLHTGTATQLPLVPLALP-----DSVLGKDTTHAVQGVVYRG- -L 202  
 Db 200 LIGRTAEQVMEIGHALIEHIPAMEIRGDLVTGLPKTILMSNEIQDMRSLHLIL 259  
 QY 203 FVIRAMTAQCKELG---CRCAAVITG-----GLSRFSEVDFPPIDAQLTISGLA 251  
 Db 260 EAIRATLEDCCPELGGDVRGVTITGGALLNGIKEWLTETEEIVVPHVAQNPLESVA 317

## RESULT 5

COBG\_PSEDE STANDARD; PRT; 459 AA.  
 ID COBG\_PSEDE STANDARD; PRT; 459 AA.

AC P21637;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE COBG protein (EC 1.1.1.1).  
 GN COBG.  
 OS Pseudomonas denitrificans.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=43306;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=SC510;  
 RX MEDLINE=91008976; PubMed=2211521;  
 RA Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,  
 RA Blanche F., Thibaut D., Debussche L.;  
 RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas  
 RT denitrificans fragment carrying eight genes involved in  
 RT denitrification of precorrin-2 to cobyrinic acid.";  
 RL J. Bacteriol. 172:5980-5990(1990).  
 CC -1- FUNCTION: CATALYZES THE ELIMINATION OF C-20 IN PRECORRIN-6W TO  
 CC FORM PRECORRIN-6X.  
 CC -1- PATHWAY: Cobalamin biosynthesis.  
 CC -1- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN  
 CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND  
 CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M59301; AAA25795.1;  
 DR PIR: B36145; B36145.  
 DR InterPro: IPR000660; Nlr\_Sir.  
 DR InterPro: IPR005117; Nlr\_Sir\_fer.  
 DR Pfam: PF01077; Nlr\_Sir; 1.  
 DR Pfam: PF03460; Nlr\_Sir\_fer; 1.  
 DR PRINTS: PR00397; SIROHAEM.  
 DR PROSITE: PS00365; Nlr\_Sir; 1.  
 KW Cobalamin biosynthesis; Forphylin biosynthesis; Oxidoreductase; Heme;  
 KW Iron-sulfur.  
 FT METAL 338 338 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 344 344 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 381 381 IRON-SULFUR (4FE-4S) AND SIROHEME  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 459 AA; 46690 MW; EC101F4C1FDC7209 CRC64;

Query Match 6.8%; Score 93.5; DB 1; Length 459;  
 Best Local Similarity 22.08; Pred. No. 1.8;  
 Matches 78; Conservative 39; Mismatches 115; Indels 123; Gaps 18;

QY 36 RKTODEYSL-----LIHALCERAGVG-----RASLRDAFISSV-VPVLTKTITADA----- 79  
 Db 45 RPTDSDTLTPKVIATATAAERFGNIIIEITARGNLQRLGLSAASVPRLAQAIGDAEIAIA 104  
 QY 80 -----VAQISGVQPVVFG-----PWAYE-----HLPRVPEPVRAEI--GTDLVANAV 120  
 Db 105 EGLAIEVPPPLAGIDPDEIADPRPATELREALDVQVPLKLPKLSVVIDSGGRFGLGAV 164  
 QY 121 AAYVHRSACVVDGCTALT-----TFTAVDGTGLIOGVAIAPGLRTAVQSLHT-GT----- 170  
 Db 165 VADIRLQAVSTVAGVAVWVLGGTSTKASSVGTLAGNAVVPALITILEKIASLIGTTMRGR 224  
 QY 171 -----RQLPLVP--LALP-----DSVLGKDTTHA-VQAGVVRGTLF 203  
 Db 225 DLDPSEIRALCRCTSSERPAAPRSAAIPGIIHALGNADTVLGLGLAFAQVAAALASYLH 284  
 QY 204 VIRAM-----TAQCKELGC 218  
 Db 285 QVQALGANAIKLAPGHAFVGLGCPETAQAQSLAASHGFRTAEODPRNATATCAGSKG 344  
 QY 219 RCAAIVITGGLS-RLFSSEVDPPIDAQLT--LSGLAH-IARLVPTSLPLPFAVSG 269  
 Db 345 ASAWMETKGAERLIVETAPEL--LDGSLTVHLSCGCAKCARPKPSELTLVGAPSG 397

## RESULT 6

MURB\_BORPE STANDARD; PRT; 351 AA.  
 ID MURB\_BORPE STANDARD; PRT; 351 AA.  
 AC O9X6Y8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158) (UDP-N-  
 DE acetylmuramate dehydrogenase).  
 GN MURB.  
 OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Tohama I;  
 RA Pradel E.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) -> UDP-N-  
 CC acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 CC -1- PATHWAY: Peptidoglycan biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AJ238308; CAB41011.1;  
 DR HSSP: P08373; 2MBR.  
 DR InterPro: IPR003170; MurB.  
 DR Pfam: PF02873; MurB\_C; 1.  
 KW Peptidoglycan synthesis; Cell wall; Cell division; Oxidoreductase;  
 KW NADP; Flavoprotein; FAD.  
 SQ SEQUENCE 351 AA; 37364 MW; 81E10E9A148B70F2 CRC64;

Query Match 6.7%; Score 92.5; DB 1; Length 351;  
 Best Local Similarity 24.08; Pred. No. 1.6;  
 Matches 68; Conservative 36; Mismatches 94; Indels 85; Gaps 15;

QY 31 LAPDARKTODEYSLIHALCERAGVGVRSLRDAFIS-----SVVPVLTKTITADAIVAQISGV 86



```

Db 11 VAPLAPOAQD-----LRCFT-LGLASHAFVALTEPSOLPALIS-ALAPFRQL----- 58
QY 87 QVVEGPNWAYEHL-----VRIPEVRAEIGTDLVANAAVAV-----HPR 127
Db 59 --VILGGSNVPLPASIDGLVAQVRLP-----GVLVGQCADAAWVEAAAGENWHGFV 109
QY 128 SACVVVDGCTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTATQALPLVPLALPSVLGK 187
Db 110 TAC--VDNGW-----DG---LENLALIPGTVGAAPVQNIAGYVELA----- 146
QY 188 DTHAVQAGVVRGTLFLVIRAMIAQO-----KELGCRCAAVITGGLSRLFSSEVD 237
Db 147 DREHSLTAWDKGGWV-EMGAECRFAYRDSFFKHQEPGAVVIGSVRFALPRPQVLD 205
QY 238 FPPIDAQTLGLSLAHARLVPTSL-----LPPATVSGSSON 273
Db 206 YPDLQRHAALDGAAPTARAVYDAVCAIRRAKLDPDPVAVGNAGS 248

```

## RESULT 7

```

GATA_MYCLE
ID GATA_MYCLE STANDARD; PRT; 497 AA.
AC O33105;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
DE subunit A).
GN GATA OR ML1702 OR MLCB637.13.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltingworth T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jacobs K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RA Nature 409:1007-1011(2001).
CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-tRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-
CC tRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-tRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-tRNA(GLN) (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine - ADP
CC + phosphate + L-glutaminyl-tRNA(Gln) + L-glutamate.
CC -!- SUBUNIT: HETEROOLIGOMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z99263; CAB16428.1;
CC EMBL; AL583923; CAC30655.1;
CC Leproma; ML1702;
CC InterPro; IPR000120; Amidase.
CC InterPro; IPR004412; Gata.
CC Pfam; PF01425; Amidase; 1.

```

```

DR TIGRFAMS; TIGR00132; gata; 1.
DR PROSITE; PS00571; AMIDASES; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 497 AA; 51536 MW; D3723D871518BDC7 CRC64;
Query Match 6.6%; Score 91.5; DB 1; Length 497;
Best Local Similarity 19.2%; Pred. NO. 2.9;
Matches 66; Conservative 37; Mismatches 108; Indels 133; Gaps 16;
QY 6 VGNSHVFGIOGEN-GGRVCVRELFR-----APDARKTQDEYSLLIHALCER-----A 53
Db 2 IGLSHEIIRLDASTLAAKIVAKELSSVEITQACLDQIEATDDTYRAFLHVAEKALSA 61
QY 54 GVGRA-----SLRDAFI-----SSV 68
Db 62 AVDKAAAGQLSSITLAGVPLALKOVFTTMDPTTCGSKILQGWHSFYDATVTRRAAG 121
QY 69 VPLVKTIAADAVAQISGVQPVVG-----PWAYEHL----- 100
Db 122 IPILGKTNMDEFAMGSSSTENSAYGPTRNPNVDRVPGGGGGGSAALAAFAQPLAIGSDT 181
QY 101 ---VRIPEVRAEIGTDLVANAAVYHFRSACV-VVDCGTALTFTAVDGTGLIOGVAIA 156
Db 182 GGSIRQPAALTATVGKPYGVTSRY--GLVACASSLDQGGPCARTVLD-TALLHAVIAG 238
QY 157 PGLRTAVQSLHTGTATQALPLVPLALPDSVLG-----KDTTHAYOAGVVR-----GT 201
Db 239 HDARST-----VEAAVPD-IVGAAGAGESDGLHGVGVVRLRGEGYQGV 286
QY 202 LFVIRAMIAQOKEKELGCRCAAVITGGLSRLFSSEVDFFPPIDAQL 245
Db 287 LASFQAARVQ-----LTALGAT---VSEVDCPHFDYAL 316

```

## RESULT 8

```

MURI_DEIRA
ID MURI_DEIRA STANDARD; PRT; 290 AA.
AC O9R010;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamate racemase (EC 5.1.1.3).
GN MURI OR DR1586.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.J., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RA radiodurans R1.";
RA Science 286:1571-1577(1999).
CC -!- FUNCTION: PROVIDES THE (R)-GLUTAMIC ACID REQUIRED FOR CELL WALL
CC BIOSYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: L-glutamate - D-glutamate.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE ASPARTATE/GLUTAMATE RACEMASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL; L02544; AAA23089.1; -  
 HSSP; P07986; 1EXG.  
 InterPro; IPR001919; Bac\_cellose-bind.  
 InterPro; IPR003961; FN\_III.  
 InterPro; IPR003962; FNIII\_repeat.  
 InterPro; IPR001547; GH\_5.  
 Pfam; PF00041; fn3; 2.  
 Pfam; PF00150; cellulase; 1.  
 Pfam; PF00553; CBM\_2; 1.  
 PRINTS; PR00014; FNTYPEIII.  
 SMART; SM00060; FN3; 2.  
 PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 FT SIGNAL 1 39 POTENTIAL  
 FT CHAIN 40 747 ENDOGLUCANASE D.  
 FT DOMAIN 448 542 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 546 639 FIBRONECTIN TYPE-III 2.  
 ACT\_SITE 208 208 PROTON DONOR (BY SIMILARITY).  
 ACT\_SITE 349 349 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 747 AA; 78936 MW; BD15473C9D8B42BD CRC64;

Query Match 6.5%; Score 90; DB 1; Length 747;  
 Best Local Similarity 25.4%; Pred. No. 6.1;  
 Matches 58; Conservative 21; Mismatches 91; Indels 58; Gaps 11;

CC 17 GENGRCVRCVRELFRLAPDARKTODE--YSLIIHALCERAG---VGRSLRDATISSVVPV 71  
 DB 388 GDTGG-----LLLDMMKTDWEKYSTMLEPTLWKHGKYG-----LDHQVPL 430

CC 72 --LTKTIADAVAQISGVQVFGFWAYEHL---PVRIPEPVRAEIGTDLVANAVAAVYH 125  
 DB 431 GGVGSTGTGTSISQVGGTDTTAPTAPTGLRAGTPTASTVPLTWASATDTGGSGVAGYEV 490

CC 126 FRSACVV-----VDCGTALTFT--AVDGTGLIQGVAAIAPGLRTAV-----163  
 DB 491 YRGTTLVGTTTATSVTVTGLAADSAYTFSVRKDGAGNTSAASAAYTARTAGGSDVTAP 550

CC 164 ---QSLHTCTAQLPLPLALPSVLGKDTTHAVQAG---VVRGTLFVIR 206  
 DB 551 SVPTGLTAGTPTATSVPLTWAST---DTGGSGVTGYEYRGSTLVAR 595

RESULT 11  
 VG26.BPMD2  
 ID VG26.BPMD2 STANDARD; PRT; 836 AA.  
 AC 064220;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Minor tail protein GP26.  
 GN 26  
 OS Mycobacteriophage D29.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 OC unclassified Siphoviridae.  
 OX NCBI\_TaxID=28369;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98300335; PubMed=9636706;  
 RA Ford M.E., Sarkis.G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;  
 RT "Genome structure of mycobacteriophage D29: implications for phage evolution.";  
 RL J. Mol. Biol. 279:143-164(1998).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U72662; AAB58898.1;  
 DR InterPro; IPR000385; Moaa\_NifB\_PqqE.  
 DR Pfam; PF01444; Moaa\_NifB\_PqqE; 1.  
 DR PROSITE; PS01305; MOAA\_NIFB\_PQOE; 1.  
 KW PQQ; Iron-sulfur.  
 FT METAL 28 IRON-SULFUR (POTENTIAL).

DR EMBL; AF022214; AAC18467.1; -  
 FT INIT\_MET 0 BY SIMILARITY.  
 SQ SEQUENCE 836 AA; 86572 MW; 528D12ED0FCB92E6 CRC64;

Query Match 6.4%; Score 88.5; DB 1; Length 836;  
 Best Local Similarity 23.0%; Pred. No. 9.3;  
 Matches 73; Conservative 39; Mismatches 114; Indels 91; Gaps 16;

CC 12 VFIQGENGRVRCVRELFRLAPDARKTODEYSLIIHALCERAG---VGRA-----58  
 DB 409 VLTVAETLTGALTALQAIQIOMPLTLVDSFKQLSETLVTSLGPLYLPQIGEAFGQIVGAV 468

CC 59 -SLRDATISSVVPVLTITIAVAQI--SGVQPVVFGFWAYEHL--PVRIPEPVRAEIGTD 114  
 DB 469 IQLAPTIISSLIPIAF--QTLIPAIQAQLAPSLVQIV---QAFTKLMPIVIPV-----VQ 516

CC 115 LVANAVAAVHFRSACVVDGCTALTFTAVDGTGLIQGV-----AIAPGL---RTAVQS 165  
 DB 517 IVINLAA-----VVQAGASTASFLIGISRLVGLADCVGAFAEWMGSSGVQ 567

CC 166 LHTCTAQLPLPLALPSVLGKDTTHAVQAG---VVRGTLFVIRAMIAQC---Q 213  
 DB 568 VSDPVGQLP-----GKIKSWFDDAGSWLIEAGKNVVQGLINGIGMISSAVSKA 616

CC 214 KELGCRCAAVITGGL-----SRLFSSEVDF-----PPIDAQLTSLGLAHIA 254  
 DB 617 KELASSVKNAVITGFLGIHSPSRVFAEIGQFTAEFGNGFEGFQPVIEKAKALA--AELS 674

CC 255 RLVPTSLPLPATVSGSS 271  
 DB 675 QAMESGVDPSSGILAGIS 691

RESULT 12  
 POQE.METEX  
 ID POQE.METEX STANDARD; PRT; 384 AA.  
 AC P71517;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Coenzyme PQQ synthesis protein E.  
 GN POQE  
 OS Methylobacterium extorquens.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Methylobacterium group; Methylobacterium.  
 OX NCBI\_TaxID=408;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AM1 / NCIMB 9133;  
 RX MEDLINE=97195805; PubMed=9043136;  
 RA Toyama H., Chistoserdova L., Lidstrom M.E.;  
 RT "Sequence analysis of pqq genes required for biosynthesis of  
 RT pyrroloquinoline quinone in Methylobacterium extorquens AM1 and the  
 RT purification of a biosynthetic intermediate.";  
 RL Microbiology 143:595-602(1997).  
 CC -!- FUNCTION: REQUIRED FOR COENZYME PYRROLO-QUINOLINE-QUINONE (PQQ)  
 CC BIOSYNTHESIS.  
 CC -!- SIMILARITY: BELONGS TO THE MOAA / NIFB / POQE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U72662; AAB58898.1;  
 DR InterPro; IPR000385; Moaa\_NifB\_PqqE.  
 DR Pfam; PF01444; Moaa\_NifB\_PqqE; 1.  
 DR PROSITE; PS01305; MOAA\_NIFB\_PQOE; 1.  
 KW PQQ; Iron-sulfur.  
 FT METAL 28 IRON-SULFUR (POTENTIAL).

FT METAL 32 32 IRON-SULFUR (POTENTIAL).  
 FT METAL 35 35 IRON-SULFUR (POTENTIAL).  
 SQ SEQUENCE 384 AA; 41714 MW; 28FD040E261B357F CRC64;  
 Query Match 6.4%; Score 88; DB 1; Length 384;  
 Best Local Similarity 20.9%; Pred. No. 4.2;  
 Matches 59; Conservative 39; Mismatches 109; Indels 82; Gaps 11;  
 QY 3 LIDVGNHVVFGTQ-----ENGRCVVCVRELFLPADARKTODEYSLLLHALCERAGVG 56  
 Db 112 LYDVGDLHVLQSVGGDAANAERIGGLKNAQPKQMPAARVTELGLPLTNSVHRGNH 171  
 QY 57 RASLRDAFISVVPVLTKTADAVAQISGVQPVVFGPWAYEH---LP-----VRIP 104  
 Db 172 EV---PGFIDLAVKLGAKRLEVHTQYIG-----WAYVNRALAMPDKSQVDESIRV 220  
 QY 105 EPVRAEIGTDLVANAVAA---YVHFRSACV-----VDCGKTALFTAVDGT 147  
 Db 221 EAARERLKGOLVDLVDPDYKYKPRACAGGWRKLMNVTPQGVLPCHAAETI----- 274  
 QY 148 GLIQGVAIAPGLTAVOSLHTGTAQLPLVPLPDSVLGKDTTHAVQAGVVRGTLFVIRA 207  
 Db 275 -----PGLEF-----WYVTDHALGEIWTSHGFRFAAYRGTSW-MKE 308  
 QY 208 MIAQCKEL---GCRCAAVITGSLSLFSSEVDFFPIDAQL 245  
 Db 309 PCRSDDREKDGCGCRQALATGDAANTDPACSLSPHAKM 350

## RESULT 13

OL56\_STRAT STANDARD; PRT; 3519 AA.  
 AC Q07017;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Oleandomycin polyketide synthase, modules 5 and 6.  
 GN ORFB.  
 OS Streptomyces antibioticus.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=1890;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150470; PubMed=8107683;  
 RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;  
 RT "Characterisation of a Streptomyces antibioticus gene encoding a type  
 RT I polyketide synthase which has an unusual coding sequence.";  
 RL Mol. Gen. Genet. 242:358-362(1994).  
 -|- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN  
 LACTONE RING.  
 -|- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 -|- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L09654; AAA19695.1;  
 DR HSSP; P25715; 1MLA.  
 DR InterPro; IPR001227; Acetyltransferase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR003880; Ppantennin-attach.  
 DR InterPro; IPR001031; Thioesterase.  
 DR Pfam; PF00109; ketoacyl-synt; 2.  
 DR Pfam; PF00550; pp-binding; 2.  
 DR Pfam; PF00698; Acyl-transferase; 2.  
 DR Pfam; PF00975; Thioesterase; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 2.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE; PS00606; B-KETOACYL-SYNTHASE; 2.  
 DR PROSITE; PS00075; ACP\_DOMAIN; 2.  
 KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.  
 FT DOMAIN 1 ?  
 FT DOMAIN 2 ?  
 FT DOMAIN 3 3519  
 FT DOMAIN 4 501  
 FT DOMAIN 5 569  
 FT DOMAIN 6 890  
 FT DOMAIN 7 1200  
 FT DOMAIN 8 1382  
 FT DOMAIN 9 1487  
 FT DOMAIN 10 1561  
 FT DOMAIN 11 1686  
 FT DOMAIN 12 2156  
 FT DOMAIN 13 2220  
 FT DOMAIN 14 2541  
 FT DOMAIN 15 2856  
 FT DOMAIN 16 3038  
 FT DOMAIN 17 3141  
 FT DOMAIN 18 3215  
 FT ACT\_SITE ?  
 FT BINDING ?  
 FT DOMAIN 19 3270  
 FT ACT\_SITE 210  
 FT ACT\_SITE 210  
 FT NP\_BIND 1203  
 FT BINDING 1524  
 FT ACT\_SITE 1859  
 FT ACT\_SITE 2311  
 FT NP\_BIND 2859  
 FT BINDING 3178  
 FT BINDING 3178  
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64;  
 Query Match 6.3%; Score 87.5; DB 1; Length 3519;  
 Best Local Similarity 22.9%; Pred. No. 58;  
 Matches 67; Conservative 34; Mismatches 114; Indels 77; Gaps 12;  
 QY 13 FGIQGGRCVRCVRELFLPADARKTODEYSLLLHALCERAGVGRASLRDAFISVVPVL 72  
 Db 451 FGVSGTN-AHVIVEE---AP--AEPEPEPTRVVAAGDLVVPVWVSGRDA-----RAL 497  
 QY 73 TKTIADAVAQISGVQPVVFGPW-----YEHLPVRIPEPVRAEIGTDLVANAVAAVH 125  
 Db 498 RAQAARLAHVSGVSAVDVG--WSLVATRSVFEHRAVAIGSEL-----DSMAGSLAGP-- 548  
 QY 126 FRGACVVVDCGTALTFTAVDGTGLI-----QGVAIAPGLTAVOSLHTGTAQL----- 173  
 Db 549 ---AAGGVPGVSGVAPAEGRVVFVPPGCGSQWGMAGLLDACPFAEVAEACAVL 605  
 QY 174 -PLVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCKELGCRCAAV----- 223  
 Db 606 DPVTGWSLSEVLQGRDATVLRVDVQVPALWAVVSLARTWRYGVPEPAVVGSHQGEIA 665  
 QY 224 ---ITGSLSLFSSEVDFFPIDAQLTSLGLAHIAHRLVPTSLPPATVSGSSG 272  
 Db 666 AACVAGGLS-----LADGARVVVLRRAIARIAGGG 697  
 RESULT 14  
 DDH2\_HUMAN STANDARD; PRT; 285 AA.  
 ID DDH2\_HUMAN  
 AC O95865;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NG-dimethylarginine dimethylaminohydrolase 2 (EC 3.5.3.18)  
 DE (Dimethylargininase 2) (Dimethylarginine dimethylaminohydrolase 2)  
 DE (DDAHII) (G6a).  
 GN DDH2 OR DDAH OR NG30 OR G6A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99425088; PubMed=10493931;  
 RA Leiper J.M., Santa Maria J., Chubb A., Macallister R.J., Charles I.G.,  
 RA Whitley G.S., Vallance P.



```
FT TRANSMEM 315 335 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT BINDING 435 435 HEME (BY SIMILARITY).
SQ SEQUENCE 489 AA; 53313 MW; 633F233CEFO3AD7A CRC64;

Query Match 6.2%; Score 85.5; DB 1; Length 489;
Best Local Similarity 25.3%; Pred. No. 8.9;
Matches 49; Conservative 28; Mismatches 72; Indels 45; Gaps 8;

QY 71 VLTNTIADAVAQISGVQPVVFGPWAYEHLVPRIPEVRA---EIGTDLVANAAYVHFR 127
Db 211 LLAQTPADVSTVAA-----PMPNRAITSVLGVDPDEA----- 244

QY 128 SACVVDCGTALFTAVDGTGLIQGVAIAFGLRTAVQSLHTGTQAQLPLVPLALPDSVLGK 187
Db 245 AFCRLSNQAVRITDVALSASGLISLVQGFAGFR-RLRALFTHRRDNGLLREC---TVLIGK 300

QY 188 DTHAVQAGVVRGTLFVIRAMIAQCCKELGCRCAAVITGGLSRFSSEVDFPPIDAQTL 247
Db 301 LATHAQGRSLDDELEFFFAVLLLVAGVESTAHM-----ISTLEFLADYPD---QLTL 350

QY 248 SGLAHIALRVPTSL 261
Db 351 --LAQOPDLIPSAI 362
```

Search completed: June 24, 2003, 21:51:17  
Job time : 6.74823 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 28.0284 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFGIGENG.....ARLVPTSLPLPATVSSGSSN 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	1379	100.0	273	AAU91157	Treponema pallidum
2	371.5	26.9	255	AAU91154	Geobacter sulfurre
3	354	25.7	256	AAU91175	Pantothenate kinas
4	344.5	25.0	259	ABB47661	Listeria monocytog
5	342.5	24.8	258	AAU91172	Pantothenate kinas
6	342.5	24.8	262	AAU91170	Pantothenate kinas
7	337.5	24.5	260	AAU91173	Pantothenate kinas
8	327.5	23.7	258	AAU01243	B. subtilis novel
9	327.5	23.7	258	AAU91149	Bacillus subtilis
10	320.5	23.2	254	AAU91171	Pantothenate kinas

11	310	22.5	219	23	AAU91176
12	293.5	21.3	258	23	AAU91153
13	291.5	21.1	265	23	AAU91151
14	288	20.9	233	23	AAU91163
15	285	20.7	246	23	AAU91156
16	282.5	20.5	250	23	AAU91150
17	254	18.4	272	22	AAU91225
18	254	18.4	272	23	AAU91152
19	248	18.0	262	23	AAU91155
20	197.5	14.3	257	23	AAU91174
21	192.5	14.0	212	23	AAU91177
22	172.5	12.5	244	23	AAU91168
23	170.5	12.4	241	23	AAU91179
24	170	12.3	262	23	AAU91158
25	148	10.7	389	21	AAU74909
26	147.5	10.7	455	21	AAU74910
27	147.5	10.7	592	20	AAU38615
28	147.5	10.7	592	20	AAU38616
29	147.5	10.7	592	21	AAU74912
30	147.5	10.7	592	21	AAU74913
31	147.5	10.7	592	23	AAU91166
32	147.5	10.7	592	23	AAU91169
33	146.5	10.6	455	20	AAU38617
34	146.5	10.6	455	21	AAU74908
35	146.5	10.6	460	23	AAU91167
36	146.5	10.6	592	20	AAU38618
37	146.5	10.6	592	21	AAU74911
38	136.5	9.9	257	23	AAU91160
39	136	9.9	248	23	AAU91164
40	132.5	9.5	249	23	AAU91182
41	122.5	8.9	229	23	AAU91159
42	118	8.6	249	23	AAU91178
43	114.5	8.3	189	20	AAU38614
44	104.5	7.6	5532	21	AAU23752
45	104.5	7.6	5532	22	AAU65267

## ALIGNMENTS

RESULT 1  
AAU91157  
ID AAU91157 standard; Protein; 273 AA.

AC AAU91157;

DT 05-JUN-2002 (first entry)

DE Treponema pallidum pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Treponema pallidum.

PN WO200216601-A2;

XX 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

PT Identifying potential antibiotic or antimicrobial agent, comprises contacting composition comprising pantothenate kinase (Coax) protein with test compound and identifying inhibitor of the Coax protein -

XX PS Claim 6; Page 75-76; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic

XX CC comprising contacting an assay composition comprising a pantothenate

XX CC kinase (Coax) protein with a test compound, and determining the ability

XX CC of the test compound to inhibit the activity of the Coax protein, an

XX CC essential enzyme for the production of coenzyme A. Coax protein is a

XX CC valuable target for identifying bactericidal compounds. Coax modulating

XX CC agents can be used in an infectious animal model to determine the

XX CC efficacy, toxicity, or side effects of treatment with such an agent. This

XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein

XX CC described in the invention.

XX SQ Sequence 273 AA;

Query Match 100.0%; Score 1379; DB 23; Length 273;

Best Local Similarity 100.0%; Pred. No. 1.2e-143;

Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

~Qy 1 MLLIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

1 MLLIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Qy 61 RDAFISVVPLTKTIADAVAQISGVQPVVFGPWAYEHLPRIPPEVRAEIGTDLVANAV 120

Db 61 RDAFISVVPLTKTIADAVAQISGVQPVVFGPWAYEHLPRIPPEVRAEIGTDLVANAV 120

Qy 121 AAYVHFSACVVDGCTALTFTAVDGTGLIOGVAIAPGLRTAVQSLHTGTQALPLVPLAL 180

Db 121 AAYVHFSACVVDGCTALTFTAVDGTGLIOGVAIAPGLRTAVQSLHTGTQALPLVPLAL 180

Qy 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCCKELGCRCAAVITGGLSLRFSSEVDPPP 240

Db 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCCKELGCRCAAVITGGLSLRFSSEVDPPP 240

Qy 241 IDAQLTSLGLAHRLVPTSLPPATVSGSGN 273

Db 241 IDAQLTSLGLAHRLVPTSLPPATVSGSGN 273

RESULT 2

AAU91154

ID AAU91154 standard; Protein; 255 AA.

XX AC AAU91154;

XX DT 05-JUN-2002 (first entry)

XX Geobacter sulfurreducens pantothenate kinase Coax.

..W Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Geobacter sulfurreducens.

XX WO200216601-A2.

PN 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

PI WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises

PT contacting composition comprising pantothenate kinase (Coax) protein

PT with test compound and identifying inhibitor of the Coax protein

XX Claim 10; Page 72-73; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic

XX CC comprising contacting an assay composition comprising a pantothenate

XX CC kinase (Coax) protein with a test compound, and determining the ability

XX CC of the test compound to inhibit the activity of the Coax protein, an

XX CC essential enzyme for the production of coenzyme A. Coax protein is a

XX CC valuable target for identifying bactericidal compounds. Coax modulating

XX CC agents can be used in an infectious animal model to determine the

XX CC efficacy, toxicity, or side effects of treatment with such an agent. This

XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein

XX CC described in the invention.

XX SQ Sequence 255 AA;

Query Match 26.9%; Score 371.5; DB 23; Length 255;

Best Local Similarity 34.7%; Pred. No. 2.1e-32;

Matches 87; Conservative 56; Mismatches 101; Indels 7; Gaps 4;

Qy 1 MLLIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

2 LLVIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 57

Qy 61 RDAFISVVPLTKTIADAVAQISGVQPVVFGPWAYEHLPRIPPEVRAEIGTDLVANAV 120

Db 58 RAVTISVVPLTKTIADAVAQISGVQPVVFGPWAYEHLPRIPPEVRAEIGTDLVANAV 115

Qy 121 AAYVHFSACVVDGCTALTFTAVDGTGLIOGVAIAPGLRTAVQSLHTGTQALPLVPLAL 180

Db 116 AGYKRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFORASKLPVRDIIR 175

Qy 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCCKELGCRCAAVITGGLSLRFSSEVDPPP 239

Db 176 PSARIARTVNSMQAGIYGVVGLVDEIVTRKAEKSKDAPRVATGGLASLAPESKTE 235

Qy 240 PIDAQLTSLGL 250

Db 236 AVEEYLTLEGL 246

RESULT 3

AAU91175

ID AAU91175 standard; Protein; 256 AA.

XX AC AAU91175;

XX DT 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #13.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Clostridium difficile.

XX WO200216601-A2.

PN 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

PI WPI; 2002-269358/31.

XX N-PSDB; ABK34196.



PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Madueno E, De Pablos B, Wehlend J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
XX WPI; 2002-010914/01.  
XX  
XX Genomic sequence for *Listeria* monocytes, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and  
PT related polypeptides -  
XX  
XX Claim 6; SEQ ID No 366; 192pp; French.  
PS  
XX The present invention relates to the genome sequence of *Listeria*  
CC monocytes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in *L.*  
CC

```

CC monocytes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 259 AA;
Query Match 25.0%; Score 344.5; DB 23; Length 259;
Best Local Similarity 32.2%; Pred. No. 2.1e-29;
Matches 83; Conservative 56; Mismatches 108; Indels 11; Gaps 4;
Qy 1 MLLIDVGNSHVFGIQENGRCVCRBELFRLLADPARKTQDEYSLLIHALCERACVGRASL 60
:|||||: |: : : : : |:| : : : : :
Db 2 ILVIDVGNCTCTGVYEKQK----LLKHWMTTDRHRTSDGLTGNLTFNFFSYANLTSPDI 57

```

Db	58	QGIILISVVPIIMHAMETCMRVFNIRPLIVPGIKTGLNLKVDNP--REIGSDRIYNVAV	115
Qy	121	AAVYHFRSACVYVDCGTAFTFTADVDTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL	180
Db	116	AASEEETPVIVVDGFTATTCYIDESGVTQGGIAFGIMISTEALYNRAAKLPRVDIAE	175
Qy	181	PDSVLGRTDTHVAQGVVGRGTLFVIRAMIAQCOKELGCRCAAVITGGLSRLF---SSEVD	237
Db	176	SSQIIKSTVSSMQAGIFYGFGQCEGIIAEMKKQSNASPVVATGGLGARMIITEKSSAVD	235
Qy	238	FPPIDAQTLTSLGLAHIAH	255
Db	236	I--LDPLFLTKGLELLYR	251
RESULT	5		
AAU91172			
ID	AAU911172	standard; Protein; 258 AA.	
XX	AAU911172;		
XX			
XX	05-JUN-2002	(first entry)	
XX			
DE	Pantothenate kinase (CoaX) #10.		
XX			
KW	Pantothenate kinase; CoaX; antibiotic; antimicrobial;		
KW	pantothenate kinase modulator; coenzyme A; bactericidal compound.		
XX			
OS	Bacillus stearothermophilus.		



KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX Caulobacter crescentus.  
XX WO200216601-A2.  
XX 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26531.  
XX 24-AUG-2000; 2000US-227860P.  
XX 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
PI Yocum RR, Patterson TA;  
XX WPI; 2002-269358/31.  
XX N-PSDB; ABK54194.  
XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
PS Claim 10; Page 102-103; 128pp; English.  
XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential target for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 260 AA;  
Query Match 24.5%; Score 337.5; DB 23; Length 260;  
Best Local Similarity 32.9%; Pred. No. 1.2e-28;  
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;  
QY 1 MLLIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTODEYSLIIHALCERAGVGRASL 60  
DB 2 LLAIEQGNNTMFAL--HDGASWVAQ-WRSATESTRTADYVVMWLSQLLSMOGLGFRAI 57  
QY 61 RDAFISSVVPVLTKTADAVAQISGVQVPGVFWAYEHLVPRIPPEVRAEIGTDLVANAV 120  
DB 58 DAVIISVVPQSIPLNRLSRRYFNVEPLVTIGENAKLGIDVRIEKP--SEAGADRLVNAI 115  
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLAL 180  
DB 116 GAAVYFPGPLVVDSTATTEDIVADGAFFGIAPGINSQALHAAKLPRIARQ 175  
QY 181 P-DSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCKELGCRCAAVITGSLRFSSEVD- 237  
DB 176 PAGNRIVGTDVTSAMQSGVFGYISLIEGLVARIKAEPEMTVIATGVSALFEGSND 235  
QY 238 FPPIDAQTLTSLGAHIAH 255  
DB 236 IDHFDSDLTIRGLEIYR 253  
RESULT 8  
AAU01243  
ID AAU01243 standard; Protein; 258 AA.  
XX AAU01243;  
XX 18-JUL-2001 (first entry)  
XX

DE B. subtilis novel pantothenate kinase encoded by the gene coax.  
XX  
KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;  
KW nutritional supplement; panto-compound; pantoate.  
XX Bacillus subtilis.  
XX WO200121772-A2.  
XX 29-MAR-2001.  
XX 21-SEP-2000; 2000WO-US25993.  
XX 21-SEP-1999; 99US-0400494.  
XX 07-JUN-2000; 2000US-0210072.  
XX 28-JUL-2000; 2000US-0221836.  
XX 24-AUG-2000; 2000US-0227860.  
XX (OMNI-) OMNIGENE BIOPRODUCTS.  
PI Yocum RR, Patterson TA, Hermann T, Pero JG;  
XX WPI; 2001-218644/22.  
XX N-PSDE; AAS00984.  
XX New recombinant microorganism which overexpress a Bacillus subtilis  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate -  
XX Example 14; Fig 23; 292pp; English.  
XX The sequence represents a novel B. subtilis pantothenate kinase (encoded  
CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one Bacillus  
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.  
XX  
SQ Sequence 258 AA;  
Query Match 23.7%; Score 327.5; DB 22; Length 258;  
Best Local Similarity 29.9%; Pred. No. 1.5e-27;  
Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;  
QY 1 MLLIDVGNHVVFGIOGNGRVCVRELFRLAPDARKTODEYSLIIHALCERAGVGRASL 60  
DB 2 LLVIDVGNNTVLGVY--HDGK--LEYHWRIFSRHKTDEFGMLSLFDHSLMFEQI 57  
QY 61 RDAFISSVVPVLTKTADAVAQISGVQVPGVFWAYEHLVPRIPPEVRAEIGTDLVANAV 120  
DB 58 DGIITISSVVPPIFALERMCTKYFHIEPQIVGPGMKTGLNKKYDNP--KEVGADRIVNAV 115  
QY 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIOGVATAPGLRTAVQSLHTGTALPLVPLAL 180  
DB 116 AAHLHGNPLVVDFTATTTCYIDENKQYMGGAIAPIGTISTEALYSRAAKPRIETR 175  
QY 181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCKELGCRCAAVITGSLRFSSEVD-FP 239  
DB 175 PDNIIGKNVTSAMQSGILFYGQVEGIVKRMKQAKQDLKVIATGGLAPLIANESDCID 235  
QY 240 PIDAQTLTSLGL 250  
DB 236 IVDPPFLTKGL 246  
RESULT 9  
AAU91149



238 FPFIDAQETLSGL 250

Db 178 RPAKVGSNVVSAMQSGIYFGYIGLVEELVRRIQTELG 215

238 FPFIDAQETLSGL 250





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.95745 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453a-10  
Perfect score: 1379  
Sequence: 1 MLLIDVGNHVVFGIQENG.....ARLVPTSLPPATVSGSGN 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1379	100.0	273	2 D71326	conserved hypothet
2	345.5	25.1	259	2 AF1464	conserved hypothet
3	344.5	25.0	259	2 AF1102	conserved hypothet
4	337.5	24.5	261	2 B87489	transcription acti
5	320.5	23.2	254	2 F83660	hypothetical prote
6	316.5	23.0	273	2 E97293	probable transcrip
7	291.5	21.1	265	2 T36391	hypothetical prote
8	288	20.9	233	2 S66100	conserved hypothet
9	285	20.7	246	2 D72320	conserved hypothet
10	267	18.4	274	2 H86937	conserved hypothet
11	254	18.4	272	2 A70955	hypothetical prote
12	248	18.0	262	2 F75516	conserved hypothet
13	170	12.3	262	2 F70165	conserved hypothet
14	151.5	11.0	224	2 A95711	conserved hypothet
15	147.5	10.7	592	2 B81009	Bira protein/Bvg a
16	147.5	10.7	592	2 H82031	probable biotin-(a
17	136.5	9.9	257	2 S75559	hypothetical prote
18	136	9.9	248	2 H83111	hypothetical prote
19	132	9.6	276	2 A12292	hypothetical prote
20	122.5	8.9	229	2 E70465	hypothetical prote
21	105	7.6	380	2 D84214	NADH dehydrogenase
22	105	7.6	430	2 A87608	hypothetical prote
23	98.5	7.1	679	2 S64258	hypothetical prote
24	94	6.8	383	2 C70845	probable nagA prot
25	93.5	6.8	335	2 C69660	cell-shape determi
26	93.5	6.8	459	2 B36145	cobG protein - pse
27	93	6.7	335	2 C83755	cell-shape determi
28	93	6.7	347	2 H75427	S-layer-like array
29	92	6.7	604	2 T04351	viviparous-14 prot

30	91.5	6.6	497	2 T45406	probable amidase (
31	91.5	6.6	2512	2 E70751	probable nrp prote
32	91.5	6.6	4077	2 T17484	hypothetical prote
33	90.5	6.6	334	2 T35939	probable transport
34	90.5	6.6	369	2 C75378	glutamate racemase
35	90.5	6.6	487	2 D87546	benzaldehyde dehyd
36	90	6.5	517	2 B87644	4-coumarate-coA li
37	90	6.5	747	2 B47093	cellulase (SC 3.2.
38	39.5	6.5	309	2 G83044	lipase LipC PA4813
39	38.5	6.4	302	2 G70614	hypothetical prote
40	38.5	6.4	837	2 H72802	minor tail subunit
41	38.5	6.4	3413	2 T17467	rifamycin polyketi
42	88	6.4	242	2 A82637	conserved hypothet
43	88	6.4	268	2 A83130	conserved hypothet
44	88	6.4	359	2 AB3407	phosphoribosylform
45	87.5	6.3	3519	2 S43048	polyketide synthas

ALIGNMENTS

RESULT 1  
D71326  
conserved hypothetical protein TP0431 - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D71326  
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.;  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.;  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770; PMID:9665876  
A:Accession: D71326  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-273 <COL>  
A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:ARC65417.1; PID:g3.  
A:Experimental source: strain Nichols  
C:Genetics:  
A:Gene: TP0431  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match	100.0%	Score 1379;	DB 2;	Length 273;
Best Local Similarity	100.0%	Pred. No. 2.7e-108;		
Matches 273;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLLIDVGNHVVFGIQENGGRVVCVRELFLAPDARKTQDEYSLIIHALCERAGVGRASL	60	
Db	1	MLLIDVGNHVVFGIQENGGRVVCVRELFLAPDARKTQDEYSLIIHALCERAGVGRASL	60	
QY	61	RDATSSVVPVLTITADAVAQISGVQVVPVFGWAVEHLFVRIPEPVRAETGTDLVANAV	120	
Db	61	RDATSSVVPVLTITADAVAQISGVQVVPVFGWAVEHLFVRIPEPVRAETGTDLVANAV	120	
QY	121	AAVHFRSACVVVDCGTALTFTAVDGTGLTQGVIAIAPGLRTAVQSLHTGTAQLPLVPLAL	180	
Db	121	AAVHFRSACVVVDCGTALTFTAVDGTGLTQGVIAIAPGLRTAVQSLHTGTAQLPLVPLAL	180	
QY	181	PDVSLGKDTTHAVQAGVVRGTLFVIRAMIACQKELGCRCAAVITGSLRFSSEVDFPP	240	
Db	181	PDVSLGKDTTHAVQAGVVRGTLFVIRAMIACQKELGCRCAAVITGSLRFSSEVDFPP	240	
QY	241	IDAQLTSLGLAHARLVPTSLPPATVSGSGN	273	
Db	241	IDAQLTSLGLAHARLVPTSLPPATVSGSGN	273	

RESULT 2  
AF1464  
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip11:  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C;Accession: AF1464  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AF1464  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-259 <GLA>  
 A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:gl6412682; GSPDB:GN00178  
 A;Experimental source: strain Clp11262  
 C;Genetics:  
 A;Gene: lin0253  
 C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.18; Score 345.5; DB 2; Length 259;  
 Best Local Similarity 32.68; Pred. No. 1.3e-21;  
 Matches 84; Conservative 53; Mismatches 110; Indels 11; Gaps 4;

QY 1 MLLIDVGNHVVFGIOGNGRVCVRELFLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
 Db 2 ILVIDVGNCTGVYKEQK----LLKHWMTTDRHRTSDELGMTVLFNFESYANLTPSDI 57  
 QY 61 RDAFISSVVPVLTFTTIADAVAQISGVQVPGFWAYEHLVPRIPEPVAEIGTDLVANAV 120  
 Db 58 QGIITSSVVPPIHMETCMVRYFNIRPLIVGPGIKTGLNLKVDNP--REIGSDRIYNV 115  
 QY 121 AAYVHFSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTAQLPLVPLAL 180  
 Db 116 AASEEYGTPIVVDFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAE 175  
 QY 181 PDSVLGKDTTHAVOAGVVRGTLFVIRAMIAOCOKELGCRCAAVITGGLSRLF---SSEVD 237  
 Db 176 SSQIIGKSTVSMQAGIFYGFGQCEGIIAEMKKQSNATSPVAVATGGLARMITEKSSAVD 235  
 QY 238 FPPIDAQLTSLGLAHIAH 255  
 Db 236 I--LDPFLTLKGLLELYR 251

RESULT 3  
 AF1102 conserved hypothetical protein lmo0221 [imported] - *Listeria* monocytogenes (strain EGD-e)  
 C;Species: *Listeria* monocytogenes  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
 Accession: AF1102  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AF1102  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-259 <GLA>  
 A;Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl6409586; GSPDB:GN00177  
 A;Experimental source: strain EGD-e  
 C;Genetics:  
 A;Gene: lmo0221  
 C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.08; Score 344.5; DB 2; Length 259;  
 Best Local Similarity 32.28; Pred. No. 1.6e-21;  
 Matches 83; Conservative 56; Mismatches 108; Indels 11; Gaps 4;

QY 1 MLLIDVGNHVVFGIOGNGRVCVRELFLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Db 2 ILVIDVGNCTGVYKEQK----LLKHWMTTDRHRTSDELGMTVLFNFESYANLTPSDI 57  
 QY 61 RDAFISSVVPVLTFTTIADAVAQISGVQVPGFWAYEHLVPRIPEPVAEIGTDLVANAV 120  
 Db 58 QGIITSSVVPPIHMETCMVRYFNIRPLIVGPGIKTGLNLKVDNP--REIGSDRIYNV 115  
 QY 121 AAYVHFSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTAQLPLVPLAL 180  
 Db 116 AASEEYGTPIVVDFGTATTCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAE 175  
 QY 181 PDSVLGKDTTHAVOAGVVRGTLFVIRAMIAOCOKELGCRCAAVITGGLSRLF---SSEVD 237  
 Db 176 SSQIIGKSTVSMQAGIFYGFGQCEGIIAEMKKQSNATSPVAVATGGLARMITEKSSAVD 235  
 QY 238 FPPIDAQLTSLGLAHIAH 255  
 Db 236 I--LDPFLTLKGLLELYR 251

RESULT 4  
 B87489 transcription activator, probable Baf family [imported] - *Caulobacter crescentus*  
 C;Species: *Caulobacter crescentus*  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C;Accession: B87489  
 R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kohn, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A;Reference number: B87249; MUID:21173698; PMID:11259647  
 A;Accession: B87489  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-261 <STO>  
 A;Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148  
 C;Genetics:  
 A;Gene: CC1935  
 C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.58; Score 337.5; DB 2; Length 261;  
 Best Local Similarity 32.98; Pred. No. 6.1e-21;  
 Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;

QY 1 MLLIDVGNHVVFGIOGNGRVCVRELFLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
 Db 3 LLAIEQGNNTMPAI---HDGASVVAQ--WRSATSTRTADDEVVYVWLSQLLSMOGLGFRAI 58  
 QY 61 RDAFISSVVPVLTFTTIADAVAQISGVQVPGFWAYEHLVPRIPEPVAEIGTDLVANAV 120  
 Db 59 DAVIISVVPQSIENLRLNRRYFNVEPLVIGENAKLGIDVRIEKP--SEAGADRLVNAI 116  
 QY 121 AAYVHFSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTAQLPLVPLAL 180  
 Db 117 GAAVYGPVLVIDSGTATTFDIYVADGAFEGGIAPGINLSMQALHRAAKLPRIAIQR 176  
 QY 181 P--DSVLGKDTTHAVOAGVVRGTLFVIRAMIAOCOKELGCRCAAVITGGLSRLFSSEVD- 237  
 Db 177 PAGNRIVGTDVTSAMQSSVFWGYISLIEGLVARIKAEERGEPTVIATGVSALFEGATDS 236  
 QY 238 FPPIDAQLTSLGLAHIAH 255  
 Db 237 IDHFDSDLTIRGLLEIYR 254

RESULT 5  
 F83660 hypothetical protein BH0086 [imported] - *Bacillus halodurans* (strain C-125)  
 C;Species: *Bacillus halodurans*  
 C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C;Accession: F83660  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;

Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: F83660  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <STO>  
A:Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN000  
A:Experimental source: strain C-125  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 23.2%; Score 320.5; DB 2; Length 254;  
Best Local Similarity 31.4%; Pred. NO. 1.6e-19;  
Matches 80; Conservative 57; Mismatches 103; Indels 15; Gaps 5;  
QY 1 MLLIDVGNHVVFGIOGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
2 ILVIDVGNHVVFGIOGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 57  
QY 61 RDAFISVVPVLTNTIADAVAGISGVQVPGWYAEHLVPRIPVRAEIGTDLVANAV 120  
DB 58 DGIIVSVVPPMFSLEOMCKYFHTVPIIGPGIKTGLNKKYDNP--KEVGADRIYNAV 115  
QY 121 AAYVHRSACVVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTQALPLVPLAL 180  
DB 116 AAIELYGPVAVDFGTATTCLINEKKOYAGGVAPGIMISTEALYHRASKLPRIETAK 175  
QY 181 PDSVLGKDTTHAVQGVVGRGLF---VIRAMIAQCQKELGCRCAAVITGSLRFSSEV 236  
DB 176 PRQVGTNTIDSMGSIPIGVYQVQDVGVKMKQAQSE----PKVIATGGLAKLIGTES 231  
QY 237 D-FPPIDAQLTSLGL 250  
DB 232 ETIDVIDSFLIKGL 246

RESULT 6  
E97293  
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
A:Accession: E97293  
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97293  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-273 <KUR>  
A:Cross-references: GB:AB001437; PIDN:AAK01136.1; PID:g15026270; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3200  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 23.0%; Score 316.5; DB 2; Length 273;  
Best Local Similarity 30.6%; Pred. NO. 3.7e-19;  
Matches 78; Conservative 59; Mismatches 107; Indels 11; Gaps 4;  
QY 1 MLLIDVGNHVVFGIOGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60  
DB 13 ILVIDVGNHVVFGIOGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 68  
QY 61 RDAFISVVPVLTNTIADAVAGISGVQVPGWYAEHLVPRIPVRAEIGTDLVANAV 120  
DB 69 EGVLISSVVPNIMVSLHMKRYKFNPLVVGPGIKTGLNKKYDNP--KEVGADRIYNAV 126  
QY 121 AAYVHRSACVVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTQALPLVPLAL 180

DB 127 AAHEYIKRSLIIDFGTATTTCAVRENGDYLGAICPGIKVSSEALFEKAALPRVELIK 186  
QY 181 PDSVLGKDTTHAVQGVVGRGLF---FVIRAMIAQCQKELGCRCAAVITGSLRFSSEV 236  
DB 187 PAYAICKNTISSIOSGIVVYIGVYIVVERKKEELQEEGEKEPLVATGGLAKLISEEA 246  
QY 237 -DFFPIDAQLTSLGL 250  
DB 247 KNVDVINDPFLTEGL 261

RESULT 7  
T36391  
hypothetical protein SCE94.31c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
A:Accession: T36391  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.  
submitted to the EMBL Data Library, April 1999  
A:Reference number: 221573  
A:Accession: T36391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-265 <OLI>  
A:Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOPDB:SCE94.31c  
A:Experimental source: strain A3(2)  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 21.1%; Score 291.5; DB 2; Length 265;  
Best Local Similarity 33.1%; Pred. NO. 4.5e-17;  
Matches 89; Conservative 47; Mismatches 108; Indels 25; Gaps 9;  
QY 1 MLLIDVGNHVVFGI-QGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 59  
DB 2 LLTIDVGNHVVFGI-QGNGRCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRAS 56  
QY 60 LRDAF---ISSVVPVLTNTIADAVAGISG-VQPVVFGWYAEHLVPRIPVRAEIGTD 114  
DB 57 LGDGDGTGTAICATVPSVLHLEHREYVYGVDPVAVLVEGVKGTGVPILTDPH--KEVGAD 114  
QY 115 LVANAVAAVYHFRSACVVDCGTALTFTAVDGTGLIOGVAIAPGLRTAVOSLHTGTQALP 174  
DB 115 RIINAAVELYGGPAIVDFGTATTFDAVSARGEYIGVYIAPGIEISVEALGVKQAQLR 174  
QY 175 LVPLALPSVLGKDTTHAVQGVVGRGLFVIRAMIAQCQKELG---CRCAAVITGSLRL 231  
DB 175 KIEVARPRSVIGKNTVEAMQSGIVGFGAGQVGVVNRMARLADDDPDVTVIATGGLAPM 234  
QY 232 F---SSEVDFPPIIDAQLTSLGLAHIAIRLV 257  
DB 235 VLGESSVID--EHEPWLITMGL-----RLV 257

RESULT 8  
S66100  
conserved hypothetical protein yacB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000  
A:Accession: S66100; E69740  
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994  
A:Title: Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* c  
A:Reference number: S65967; MUID:96051385; PMID:7584024  
A:Accession: S66100  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <OGA>  
A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BA05305.1; PID:d1005847; PID:g46  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199  
R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; B  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E69740  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <KUN>  
A:Cross-references: GB:299104; GB:AL009126; NID:q2632267; PIDN:CAB11846.1; PID:ell82003;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yacB  
A:Start codon: TTG  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 20.98; Score 288; DB 2; Length 233;  
Best Local Similarity 29.78; Pred. No. 7.6e-17;  
Matches 62; Conservative 56; Mismatches 85; Indels 6; Gaps 3;  
QY 1 LMLIDVGNHSHVFGIOGNGRVCVRELFLAPDARKTQDEYSLIIHALCERAGVGRASL 60  
DB 2 LLVIDGNTTVLGYY--HDGK--LEYHWRIETSRHKTDEFGMLIRSLDFHSGLMPEQI 57  
QY 61 RDAFISVVPVLTAKTIADAVAQISGVQPVVFGPWAYEHLVPRIPEPVRAEIGTDLVANV 120  
DB 58 DGIISVVPVPIPFALERMCTKVFHIEPQIVGPMGTGLNKKYDNP--KEVGADRIYNAV 115  
QY 121 AAYVHFRSACVVDCGTALTFTAVDGTGLIQGVIAIAPGLTAVQSLHTGTGAQLPLVPLAL 180  
DB 116 AAHLXGNPLVDFGATTCTYIDENKQYMGAAFGITISTEALYSRAAKLPRIETR 175  
QY 181 PSDVLGKDTTHAVQAGVVRGTLFVIRAMI 209  
DB 176 PDNIGKNTYSAMQSGILFVYVQVEGIV 204

RESULT 9  
D72320  
Conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
Species: Thermotoga maritima  
Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
Accession: D72320  
Reference number: A72200; MUID:99287316; PMID:10360571  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-246 <ARN>  
A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TW0883  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 20.78; Score 285; DB 2; Length 246;  
Best Local Similarity 32.48; Pred. No. 1.5e-16;  
Matches 84; Conservative 43; Mismatches 104; Indels 28; Gaps 9;  
QY 2 LLIDVGNHSHVFGIOGNGRVCVRELFLAPDARKTQDE-----YSLIIHALCERAGV 56

DB 3 LLVDVGNHSHVFGIOGNGRVCVRELFLAPDARKTQDEYSLIIHALCERAGVGRASL 58  
QY 57 RASLRDAFISVVPVLTAKTIADAVAQISGVQPVVFGPWAYEHLVPRIPEPVRAEIGT 113  
DB 59 -----VASVVPVTPQNTVIERFSOKVFHISPI-----WVKAKNGCVKWNKNP--SEVGA 104  
QY 114 DLVANAVAVHFRSACVVDCGTALTFTAVDGTGLIQGVIAIAPGLTAVQSLHTGTGAQL 173  
DB 105 DRVANVAVAFKEYKNGKIIDMGATTVDLV-VNGSYEGGAILPFFGFMVHSLFRGTAKL 163  
QY 174 PLVPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOCOKELGCRCAAVITGGSLRFLS 233  
DB 164 PLVEVKPADFVVGKDTFENTRLGVVNGSVYALEGIIGRI-KEYVGDLPVLTGGQSKIVK 222  
QY 234 SEVDFFPIDAQLTSLGLAH 252  
DB 223 DMKHEIFDEDLTIKGVYH 241

RESULT 10  
H86937  
Conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: H86937  
R: Cole, S.T.; Eiglmeter, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.  
R: Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holr  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: H86937  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <SPO>  
A:Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: ML0232  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 19.48; Score 267; DB 2; Length 274;  
Best Local Similarity 31.78; Pred. No. 5.3e-15;  
Matches 85; Conservative 43; Mismatches 118; Indels 22; Gaps 7;  
QY 1 LMLIDVGNHSHVFGIOGNGRVCVRELFLAPDARKTQDEYSLIIHALCERAGVGRASL 60  
DB 2 LLVIDGNTTVLGYY--HDGK--LEYHWRIETSRHKTDEFGMLIRSLDFHSGLMPEQI 58  
QY 61 RDAFISVVPVLTAKTIADAVAQISGVQPVVFGPWAYEHLVPRIPEPVRAEIGTDLVANV 119  
DB 59 AGAALSTVPSVLHEVRIMLDQWPSVPHVLIETGVTGIPLLDNP--KEVGADRIYVNC 116  
QY 120 VAAVHFRSACVVDCGTALTFTAVDGTGLIQGVIAIAPGLTAVQSLHTGTGAQLPLVPLA 179  
DB 117 LAAPHKFGQAAIIVDFGSSICVDVVSAGKGFELGAIAPGVVSSDAARAARVLA 176  
QY 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAOCOK-----ELGCRCAAVITGGSLRFL 232  
DB 177 RPSVVGKNTVECNQAGVFGFAGLVGDLGRMQRDVEEFSGLGNRVAVVATGHTAPLL 236  
QY 233 SSE---VDFPPIDAQLTSLGLAHIAVL 257  
DB 237 LPELHTVDH--YDRHLTLHGL----RLV 258

RESULT 11  
A70955  
Hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70955

R.; Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gencies, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID: 98295987; PMID: 9634230  
A: Accession: A70955  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-272 <COL>  
A: Cross-references: GB:295557; GB:AL123456; NID:93242276; PIDN: CAB08944.1; PID: g2113976  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: RV3600c  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 18.4%; Score 254; DB 2; Length 272;  
Best Local Similarity 31.3%; Pred. No. 6.5e-14;  
Matches 84; Conservative 42; Mismatches 118; Indels 24; Gaps 7;  
QY 1 MLLIDVGNHVVFGIQQGNGRVCVRLFRAPDARKTQDEYSLIHALCERAGVGRASL 60  
DB 2 LLAIDVRNTHVVGSLGKHEKAKVQVQWRINTESEVTADELALIDGL-----IGEDSE 56  
QY 61 R---DAFISVVVPLTKTIADAVAQISGVQVVPVFGPWAYEHLFVRIPEPVRAEIGTDLVA 117  
DB 57 RLGTAAALSTVPSVLEHVRIMLDQYWPSPVPHVLIBEGVVRTGIPLLVNDP--KEVGADRV 114  
QY 118 NAVAAYVHFRSACVVDGCTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTCTAQLPLVP 177  
DB 115 NCLAAVDYRFRKAAIYVDFGSSICVDWSAKGEFLGAGIAPGVQVSSDAAAARSALRRVE 174  
QY 178 LALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELG-----CRCAAVITGGLSRLE 232  
DB 175 LARPSVGVKNVCEQMAGAVFGAGLVGLVGRIDREDSVGSVDHDAIVATGHTAPLL 234  
QY 233 SSE-----VDFPPIDAQLTSLGLAHIARLV 257  
DB 235 LPELHTVDH--YDQHLTLQGL----RLV 256

RESULT 12  
E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C: Species: Deinococcus radiodurans  
C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C: Accession: E75516  
R.; White, O.; Eisen, J. A.; Heidelberg, J. F.; Hickey, E. K.; Peterson, J. D.; Dodson, R. J.; M.; Shen, M.; Vamathevan, J. J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M. S.; Smith, H. O.; Venter, J. C.; Fraser, C. M. Science 286, 1571-1577, 1999  
A: Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A: Reference number: A75250; MUID: 20036996; PMID: 10567266  
A: Accession: E75516  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-262 <WHI>  
A: Cross-references: GB:AE0001905; GB:AE000513; NID: g6458144; PIDN: AAF10040.1; PID: g645814  
A: Experimental source: strain R1  
C: Genetics:  
A: Gene: DR0461  
A: Map position: 1  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c  
Query Match 18.0%; Score 248; DB 2; Length 262;  
Best Local Similarity 28.7%; Pred. No. 2e-13;  
Matches 75; Conservative 49; Mismatches 117; Indels 20; Gaps 8;  
QY 1 MLLIDVGNHVVFGIQQGNGRVCVRLFRAPDARKTQDEYSLIHALCERAG--VGPA 58  
DB 6 LLAVDIGNTTVTGLADAGS---ALHTWRINTEMLPDDIALQLHGLFTLAGAPIRA 62

QY 59 SLRDAFISSVVPVLTKTITADAVAQISGVQVVPVFGPWAYEHLFVRIPEPVRAEIGTDLVA 117  
DB 63 ----AVLSSVAPVPGENYALAKRHFMDAFV---SAENLPDVTVELDTPGSGVGRDLCL 115  
QY 118 NAVAAYVHFRSACVVDGCTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTCTAQLPL 175  
DB 116 NLFGAEKYLGGDIYAVVDFGISTNFDVVGRRFLGGILATGAQVSADALFARAALKPR 175  
QY 176 VPLALPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLE---F 232  
DB 176 ITLOAPETAIGNKTNVHALQSLGVFGVAEMVDGLLRIRAEPLCEAVAVATGGFSRTVQGI 235  
QY 233 SSEVDFPPIDAQLTSLGLAHIA 253  
DB 236 COEIDY--YDETITLRLGLVEL 254  
RESULT 13  
F70165  
conserved hypothetical protein BB0527 - Lyme disease spirochete  
C: Species: Borrelia burgdorferi (Lyme disease spirochete)  
C: Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C: Accession: F70165  
R.; Fraser, C. M.; Castjens, S.; Huang, W. M.; Sutton, G. G.; Clayton, R.; Lathigra, R.; son, D.; Peterson, J.; Kerlavage, A. R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M. D.; Horst, K.; Roberts, K.; Hatch, Nature 390, 580-586, 1997  
A: Authors: Smith, H. O.; Venter, J. C.  
A: Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A: Reference number: A70100; MUID: 98065943; PMID: 9403685  
A: Accession: F70165  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-262 <KLE>  
A: Cross-references: GB:AE001154; GB:AE000783; NID: g2688431; PIDN: AAC66882.1; PID: g2  
A: Experimental source: strain B31

Query Match 12.3%; Score 170; DB 2; Length 262;  
Best Local Similarity 25.3%; Pred. No. 7e-07;  
Matches 64; Conservative 50; Mismatches 123; Indels 16; Gaps 6;  
QY 2 LLIDVGNHVVFGIQQGNGRVCVRLFRAPDARKTQDE-YSLIHALCERAGVGRASL 60  
DB 9 LLIDIGNTSIAFALEKDNQ----VNLFIKMTNMLRYDEVYSFFSENF-----DFNV 57  
QY 61 RDAFISVVVPLTKTIADAVAQISGVQVVPVFGPWAYEHLFVRIPEPVRAEIGTDLVANAV 120  
DB 58 NKVFISVVVPIILNETKKNVIFSPFKIKPLFIFGDLNVDLTNPNYKSKDFLLGSDVFANLV 117  
QY 121 AAYVHFR-SACVVDGCTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTCTAQLPLVPLA 179  
DB 118 ALENYSFENLVVDLTGCTACTIPAVSRQDGLGINSGLPLNFNSLLDNAYLIKRPIS 177  
QY 180 LPDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLESS--EVD 237  
DB 178 TPNNLLERTSGVNSGLFYQYKYLIEGVYRDIKQMYKKFNLIITGKNADLILSLIEIE 237  
QY 238 FPPIDAQLTSLGL 250  
DB 238 F-IFNIHLTVEVG 249  
RESULT 14  
A99571  
conserved hypothetical protein. MYPD\_4730 [imported] - Mycoplasma pulmonis (strain  
C: Species: Mycoplasma pulmonis  
C: Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C: Accession: A99571  
R.; Chabaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer Nucleic Acids Res. 29, 2145-2153, 2001  
A: Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma  
A: Reference number: A99512; MUID: 21267165; PMID: 11353084  
A: Accession: A99571

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-224 <QUR>  
A:Cross-references: GB:AL445566; PID:g14089887; PIDN:CAC13646.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPV.4730  
A:Genetic code: SGC3

Query Match 11.0%; Score 151.5; DB 2; Length 224;  
Best Local Similarity 25.6%; Pred. No. 2.1e-05;  
Matches 62; Conservative 47; Mismatches 102; Indels 31; Gaps 12;  
QY 2 LLIDVGNHVVFGIOGNGRVCVRELFRAPARKTDQ-----EYSLLIHALCERAGVGR 57  
DB 3 IFIDLGNSFAKALKNEKAHF-----LFRL-----KTQSVVDPSFKSFNFLE---FNK 50  
QY 58 ASLRDAFISVVVPVLTKTAD-AVAQISGVQPVVFGWPAYEHLPLVRIPEPVRAEIGTDLV 116  
DB 51 LDVKEILICSVRNAKENQILEFKLSIFKNKIDFFIHKQSLSLVKPCQKELTSEIGLDIV 110  
Y 117 ANAAYAV-HFRSACVVVDGTTAVDTAVDGTGLIOGVATAPGLRTAVQSLHTGTQALP- 174  
DB 111 ANAY--YVLHKSNAIFISLGTAVITQIKNS-IGISYIPGIYQSFKNFFNVAKIES 167  
QY 175 -LVPLALPDSVLGKDTTHAVQAGVGRGTLEVRAMIAQCOKELGRCRAAVITGG---LSR 230  
DB 168 NFNIKIP-PILGKTTLESISLVRGSLVLLKGYIDEIK-----TSDIFITGTITFSK 222  
QY 231 LF 232  
DB 223 IF 224

## RESULT 15

B81009  
BifA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: B81009  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Olin, H.; Vanathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: B81009  
A:Status: preliminary  
Molecule type: DNA  
Residues: 1-592 <TET>  
A:Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB2075

Query Match 10.7%; Score 147.5; DB 2; Length 592;  
Best Local Similarity 25.4%; Pred. No. 0.00014;  
Matches 71; Conservative 39; Mismatches 103; Indels 67; Gaps 12;  
QY 2 LLIDVGNHVVFGIOGNGRVCVRELFRAPARKTDQDEYSLLIHALCERAGV 55  
DB 341 LLLDGGNSRLKWA-WVNGTFTAVGSAFYRDLSPGLAEWAERKADGNVRIVGCAVC--GEF 397  
QY 56 GRASLRDAFISVV--VPVLTKITADAVAQISGVQPVVFGWPAYEHLPLVRIPEPVRAEIGT 113  
DB 398 KKAQVOEQARKIEWLP-----SSAQLGIRN-----HYRH-----PE-----EHGS 434  
QY 114 DLVANAAVAVHFRSACVVVDGTTAVDTAVDGTGLIOGVATAPGLRTAVQSLHTGTQAL 173  
DB 435 DRWFNALGSRRRSRNACVVVSCGTAVTVDALTDGHLGGLTTPGPHLMKESLAVTANL 494  
QY 174 PLVPLALPDSVLGK-----DTHAVQAGVGRGTLEVRAMIAQCOKELGRCRAAVITGG 227

Db 495 -----NRHAGRRYFPPTTTGNAAVSGMMDAVCGSVMMHGRUKETGAG----- 538  
QY 228 LSRLESSEVDPPPIDAQILTSGLAHARLVPTSLIPPAIV 267  
Db 539 -----KPYDVIITGGGAAKVAEALPPAFLAENTV 567

Search completed: June 24, 2003, 22:03:08  
Job time : 11.9574 secs













```
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-28

Query Match          6.4%; Score 81.5; DB 4; Length 1220;
Best Local Similarity 21.7%; Pred. No. 4.6;
Matches 39; Conservative 31; Mismatches 85; Indels 25; Gaps 6;

QY 78 FHISPIWKAKNGCVKNNKPNSEVGADRVANVAFVKEYGKNG-----IIDMGAT 130
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 FNKDTIFDKONGAVTFDIKAPIGVNNRNLNYASFGNISVSGGNTVTFKLLASSTAQ 320
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 TVDLVNGSY---EGGAILP-----GPFMVHSLFRGT-AKLPLVEKPADFVYVK 177
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 TPGVFINSKHNAGSGSLEFRTEGSTKVGFLINNDLTLNATGNNISLQVEGIDGMIGK 380
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 DTEENIRLGVNGSVYALLEGITIGRIKEYIGDLPVVLTTGGQSKIVKDMIKHEFFDGLTIK 237
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 GVAKKNTTFAGGNI--TFGSKKAITEGNATINNANVTILGSDFDNHQ---KPLTIK 435

; RESULT 13
; US-09-206-942-26
; Sequence 26, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 1226
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-26

Query Match          6.4%; Score 81.5; DB 4; Length 1226;
Best Local Similarity 21.7%; Pred. No. 4.7;
Matches 39; Conservative 31; Mismatches 85; Indels 25; Gaps 6;

QY 78 FHISPIWKAKNGCVKNNKPNSEVGADRVANVAFVKEYGKNG-----IIDMGAT 130
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 FNKDTIFDKONGAVTFDIKAPIGVNNRNLNYASFGNISVSGGNTVTFKLLASSTAQ 326
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 TVDLVNGSY---EGGAILP-----GPFMVHSLFRGT-AKLPLVEKPADFVYVK 177
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 TPGVFINSKHNAGSGSLEFRTEGSTKVGFLINNDLTLNATGNNISLQVEGIDGMIGK 386
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 DTEENIRLGVNGSVYALLEGITIGRIKEYIGDLPVVLTTGGQSKIVKDMIKHEFFDGLTIK 237
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 387 GVAKKNTTFAGGNI--TFGSKKAITEGNATINNANVTILGSDFDNHQ---KPLTIK 441

; RESULT 14
; US-09-459-146-3
; Sequence 3, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
```

```
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,146
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,117
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Catherine R
; REGISTRATION NUMBER: 34,240
; REFERENCE/DOCKET NUMBER: 8589
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-6208
; TELEFAX: (201) 235-3500
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3165 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Endothia parasitica (Cryptosporidia)
; ORGANISM: parasitica)
; STRAIN: EP713
; US-08-459-146-3

Query Match          6.4%; Score 81.5; DB 2; Length 3165;
Best Local Similarity 20.3%; Pred. No. 20;
Matches 45; Conservative 26; Mismatches 68; Indels 83; Gaps 7;

QY 11 HSFVSTEDGKTFRRWRLSTGVFTQDELEFSLHPLHPLGDAMREIKGI-----CVAS 61
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 751 HDFDLRIETTKTCRR-----NPPEE---NLQAKLEDTARKVCSVMQYNIMIASVAF 799
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 VYPTQNTVTERFSQKYFHISP-----IWKAKNGCVKNNKPNSEVGADRVANVVA 112
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 800 LVPLFTLVYVLPYLFYLVDPDGYILLPPVLMVLTNLCYGY-----ACDAWCLRFF 851
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 FYKEYGKNGII-----IDMT 128
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 852 FYEEAGKVELHVSSEFSDPSSTLLIPTMGTRGDHVPFRFANMAVLGVKTHLLKLOT 911
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 ATTVDLVNGSVGGAILPGPFMMVHSLFRG--TAKLPLVEV 168
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 912 ATYGDLENLKKGLGSLPGLYQNHYSVLGRYKAFTPHVEL 953

; RESULT 15
; US-08-459-065-3
; Sequence 3, Application US/08459065
; Patent No. 5882642
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
```

COUNTRY: U.S.A.  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,065  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/832,117  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Roseman, Catherine R  
REGISTRATION NUMBER: 34,240  
REFERENCE/DOCKET NUMBER: 8589  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-6208  
TELEFAX: (201) 235-3500  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3165 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Endothia parasitica (Cryptonectria)  
ORGANISM: parasitica  
STRAIN: EP713  
US-08-459-065-3

Query Match 6.4%; Score 81.5; DB 2; Length 3165;  
Best Local Similarity 20.3%; Pred. No. 20;  
Matches 45; Conservative 26; Mismatches 68; Indels 83; Gaps 7;  
Qy 11 HSFSITEDGKTFRRRLSTGVFQTEDELSHLPLLDGDMREIKGI-----GVAS 61  
Db 751 HFDRLIEITKTCR-----NPPPE--NLQAKLEDARKVCSVWQYINIMIASVAF 799  
Qy 62 VVPTQNTVIERFSOKYFIHSP-----IWKAKNGCVKWNKNPSEVGADRVANYVA 112  
Db 800 LVPLYFTLYVPLQFYLVDPGDIYLLPPVLWLVWVTLNLCYGY-----ACDWCRLFF 851  
Qy 113 FVKEYGKNGII-----IDMGT 128  
b 852 FVEEAGKKELVHSSEEFSSDPSSTLLIPTMTGRGDHVPFRFANNAVLAGVKTLLKLQT 911  
Qy 129 ATTVDLVVNGSYEGAILPGPFMMVHSLFRG--TAKLPLVEV 168  
Db 912 ATYGDLENLKKGLGSLLPGLQNHYSVLRYKKAFTPHVEL 953

Search completed: June 24, 2003, 21:49:22  
Job time : 8.18642 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.2158 seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSTEDG.....HEIFDEDLTKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCRTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1281	100.0	246	9	US-09-813-453A-9
2	405.5	31.7	258	9	US-09-813-453A-2
3	399.5	31.2	262	9	US-09-813-453A-45
4	389.5	30.4	254	9	US-09-813-453A-47
5	375.5	29.3	258	9	US-09-813-453A-49
6	369.5	28.8	233	9	US-09-813-453A-17
7	361.5	28.2	255	9	US-09-813-453A-7
8	359.5	28.1	260	9	US-09-813-453A-51
9	342.5	26.7	265	9	US-09-813-453A-4
10	333.5	26.0	250	9	US-09-813-453A-3
11	321	25.1	256	9	US-09-813-453A-55
12	295.5	23.1	258	9	US-09-813-453A-6
13	286	22.3	272	9	US-09-813-453A-5
14	285	22.2	273	9	US-09-813-453A-10
15	282	22.0	272	9	US-09-712-363-276
16	279.5	21.8	219	9	US-09-813-453A-57
17	260	20.3	262	9	US-09-813-453A-8
18	229.5	17.9	241	9	US-09-813-453A-63
19	228.5	17.8	257	9	US-09-813-453A-53

20	221	17.3	212	9	US-09-813-453A-59	Sequence 59, Appl
21	203	15.8	262	9	US-09-813-453A-11	Sequence 11, Appl
22	193.5	15.1	229	9	US-09-813-453A-12	Sequence 12, Appl
23	165	12.9	223	9	US-09-895-913A-74	Sequence 14, Appl
24	165	12.9	223	9	US-09-813-453A-14	Sequence 74, Appl
25	165	12.9	223	9	US-09-813-453A-67	Sequence 67, Appl
26	162	12.6	244	9	US-09-813-453A-41	Sequence 41, Appl
27	152.5	11.9	460	9	US-09-813-453A-39	Sequence 39, Appl
28	152	11.9	267	9	US-09-813-453A-15	Sequence 15, Appl
29	150.5	11.7	592	9	US-09-813-453A-22	Sequence 22, Appl
30	146	11.4	592	9	US-09-813-453A-43	Sequence 43, Appl
31	145.5	11.4	249	9	US-09-813-453A-70	Sequence 70, Appl
32	145	11.3	248	9	US-09-813-453A-20	Sequence 20, Appl
33	139.5	10.9	209	9	US-09-813-453A-21	Sequence 21, Appl
34	138	10.8	257	9	US-09-813-453A-13	Sequence 13, Appl
35	117.5	9.2	249	9	US-09-813-453A-61	Sequence 61, Appl
36	103	8.0	242	9	US-09-813-453A-65	Sequence 65, Appl
37	97	7.6	317	12	US-10-043-238-1	Sequence 1, Appl
38	97	7.6	317	12	US-10-043-238-3	Sequence 3, Appl
39	90.5	7.1	351	9	US-10-200-910-2	Sequence 2, Appl
40	90.5	7.1	351	9	US-10-200-910-10	Sequence 10, Appl
41	90.5	7.1	436	9	US-10-076-535-2	Sequence 2, Appl
42	90.5	7.1	436	9	US-10-200-910-6	Sequence 6, Appl
43	87.5	6.8	541	10	US-09-873-409-7	Sequence 7, Appl
44	87.5	6.8	1058	10	US-09-873-409-4	Sequence 4, Appl
45	87.5	6.8	1222	10	US-09-873-409-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-9  
; Sequence 9, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 246  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
US-09-813-453A-9  
Query Match 100.0%; Score 1281; DB 9; Length 246;  
Best Local Similarity 100.0%; Pred. No. 1.1e-118;  
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MYLLVDVGNTHSVFSTEDGKTFRRRLSTGVFTQTEDELSHLPLGLGDAMREIKGIGVA 60  
Db 1 MYLLVDVGNTHSVFSTEDGKTFRRRLSTGVFTQTEDELSHLPLGLGDAMREIKGIGVA 60  
QY 61 SVVPTQNTVIERFSQKIFHISPIWVAKKNCVKNVKNPSEVGADRVANVAVFVKEYGKN 120  
Db 61 SVVPTQNTVIERFSQKIFHISPIWVAKKNCVKNVKNPSEVGADRVANVAVFVKEYGKN 120  
QY 121 GIIDMGATATVDLVVNGSYEGGAILPGPFMMVHSLFRGAKPLPLVEVKPADVVGKDT 180  
Db 121 GIIDMGATATVDLVVNGSYEGGAILPGPFMMVHSLFRGAKPLPLVEVKPADVVGKDT 180  
QY 181 ENIRLGVNGSVYALEGIIGRIKEVYGDLPVLTGGOSKIVKDMIKHEIFDEDLTKGVY 240  
|||||

```
Db 181 ENIRLGVNVSVALEGIIRIKRYVYVDLPVLTGGQSKIVKDMIKHEIFDEDLTIKGYV 240
QY 241 HFCFGD 246
Db 241 HFCFGD 246

RESULT 2
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 31.7%; Score 405.5; DB 9; Length 258;
Best Local Similarity 39.2%; Pred. No. 4.2e-32;
Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;

QY 1 MYLLVDVGNTHSVFSITEDGKTRFRWRLSTGVQFQTEDE-----LFSHLHPLLGDMRE 53
Db 1 LLLVDVGNTHSVFSITEDGKTRFRWRLSTGVQFQTEDE-----LFSHLHPLLGDMRE 56

QY 54 IKGIGVASVPTQNTVIERFSQKYEHPISPIW--IKVAKNGCVKNVK--NPSEVGADRVAN 109
Db 57 IDGIVSSVPPMPFALERCMEKFKPLVVGPGIKTGL---NIKYDNKPEVGADRVN 113

QY 110 VVAFVKEGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV 168
Db 114 AVAAIHLXGNPLVVDFTATTTCYIDENKQTMGGAIPGIIISTEALYSRAAKLPRIE 173

QY 169 KPADPVVCKDTEENIRLGVNVSVALEGIIRIKRYVYVDLPVLTGGQSKIV--KDM 224
Db 174 TRPDNIIGNTVSAMOSGILFYGVQVEGIVKRMKQAKQDLKVIATGGLIAPLIANESDC 233

QY 225 IKHEIFDEDLTIKGY 239
Db 234 I-DIVDFFLTKGL 246

RESULT 3
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-09-813-453A-45

Query Match 31.2%; Score 399.5; DB 9; Length 262;
Best Local Similarity 39.2%; Pred. No. 1.7e-31;
Matches 98; Conservative 45; Mismatches 94; Indels 13; Gaps 7;

QY 1 MYLLVDVGNTHSVFSITEDGKTRFRWRLSTGVQFQTEDELSHLHPLL---GDAMREIKGI 57
Db 1 MIFVLVDGNTNVLGVFEGBELQHRMETDRKTEDEYCYMLVKQLLEHSELSFEDVGI 60

QY 58 GVASVPTQNTVIERFSQKYEHPISPIW--KAKNGCVKNVK--NPSEVGADRVANVAF 113
Db 61 IVSSVPPMPFALERCMEKFKPLVVGPGIKTGL---NIKYPNPEVGADRVANVAG 117

QY 114 VKEYGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV 172
Db 118 IHLGSPLIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIETKPS 177

QY 173 FVVGKDTENIRLGVNVSVALEGIIRIKRYVYVDLP-VLTGGQSKIVKDMIK-HEIF 230
Db 178 SVVGRNTVSAMOSGILFYGVQVEGIVKRMKEAKQEPKVIATGGLAKLISESNVIDV 237

QY 231 DEDLTIKGYV 240
Db 238 DPFLTKGLY 247

RESULT 4
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OG2-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 30.4%; Score 389.5; DB 9; Length 254;
Best Local Similarity 35.6%; Pred. No. 1.6e-30;
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;

QY 1 MYLLVDVGNTHSVFSITEDGKTRFRWRLSTGVQFQTEDE-----LFSHLHPLLGDMRE 53
Db 1 MLIVDVGNTNVLGVYQDQETLVHWRRLATSRKTEDEYAMTVRSFLFDH---AGLOFQD 56

QY 54 IKGIGVASVPTQNTVIERFSQKYEHPISPIW--KAKNGCVKNVK--NPSEVGADRVAN 109
Db 57 IDGIVSSVPPMPFALERCMEKFKPLVVGPGIKTGL---NIKYDNKPEVGADRVN 113

QY 110 VVAFVKEGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV 168
Db 114 AVAAIHLXGNPLVVDFTATTTCYIDENKQTMGGAIPGIIISTEALYSRAAKLPRIE 173

QY 169 KPADPVVCKDTEENIRLGVNVSVALEGIIRIKRYVYVDLPVLTGGQSKIV-KDMIK 226
```



Db 174 AKPQVGTNTIDSWQSGIFGYVSQVGVKRMKAQAESEPKVIATGGLAKLIGTSET 233  
Qy 227 HEIFEDDLTKGV 239  
Db 234 IDVIDSFLTKGL 246

## RESULT 5

US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 29.3%; Score 375.5; DB 9; Length 258;  
Best Local Similarity 37.5%; Pred. No. 3.9e-29;  
Matches 94; Conservative 48; Mismatches 92; Indels 17; Gaps 9;  
Qy 1 MYLLVDGNTSVFSTEDGKTFRRWRLSTGVQFQTEDELSHLHLL---GDAMREIKGI 57  
Db 1 MIFLDVGNNTVLGVGDGLKHWRIETSRKTEDEYGMKALLNHNVGQFSDIRGI 60  
Qy 58 GVASVPTQNTVIERFSQKYPHISPIWV--KAKNGC-VKWNVKNPSEVGDADVAVVAFV 114  
Db 61 IISVVPPIMFALERMCKLYPHIKPLVPGKTKGLDKYD--NPREGADRVNAVAGI 118  
Qy 115 KEYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEVKPADF 173  
Db 119 HLYGSLIIVDFGTATTCYINERKQYMGGAIPGIMISTEALFARAALKPRIETARPDD 178  
Y 174 VVGKTEENIRLGVSVALEGIIGRIKEYVGDLP--VLTGGOSKIV---KMIKHE 228  
Db 179 IIGKNTVSAMQAGILYGVQVEGIVSRMK-AKSIKKPKVIATGGLAPLASESDII--D 235

Qy 229 IFDEDLTKGV 239

Db 236 VDPFLTLTGL 246

## RESULT 6

US-09-813-453A-17  
; Sequence 17, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-17

Query Match 28.8%; Score 369.5; DB 9; Length 233;  
Best Local Similarity 40.5%; Pred. No. 1.3e-28;  
Matches 87; Conservative 33; Mismatches 76; Indels 19; Gaps 6;  
Qy 1 MYLLVDGNTSVFSTEDGKTFRRWRLSTGVQFQTEDE-----LFSLHPLLLGDAMRE 53  
Db 1 MLFLVDGNTVTLGVYHDKLEYHWRIFSRKTEDEFGMILRSLEFDH---SGLMFEQ 56  
Qy 54 IRGIGVASVPTQNTVIERFSQKYPHISPIW--KAKNGCVKNVKNPSEVGDADVAVVAFV 109  
Db 57 IDGIISSVPPIMFALERMCKLYPHIEPQIVGPGMKTGL---NIKYDNPKEVGADRVN 113  
Qy 110 VVAFVKEYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168  
Db 114 AVAAIHLGNPLIIVDFGTATTCYIDENKQYMGGAIPGIMISTEALYSRAALKPRIE 173  
Qy 169 KPADFVVGKTEENIRLGVSVALEGIIGRIK 203  
Db 174 TRPDNIIGKNTVSAMQAGILYGVQVEGIVKRMK 208

## RESULT 7

US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 28.2%; Score 361.5; DB 9; Length 255;  
Best Local Similarity 36.8%; Pred. No. 9.2e-28;  
Matches 91; Conservative 49; Mismatches 98; Indels 9; Gaps 6;  
Qy 1 MYLLVDGNTSVFSTEDGKTFRRWRLSTGVQFQTEDE---LFSLHPLLLGDAMREIKGI 57  
Db 1 MLFLVDGNTVTLGVYHDKLEYHWRIFSRKTEDEFGMILRSLEFDH---SGLMFEQ 60  
Qy 58 GVASVPTQNTVIERFSQKYPHISPIWV--KAKNGCVKNVKNPSEVGDADVAVVAFV 115  
Db 61 IISVVPPIMFALERMCKLYPHIKPLVPGKTKGLDKYD--NPREGADRVNAVAGI 119  
Qy 116 EYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEVKPADF 174  
Db 120 KYTSLIIVDFGTATTCYINERKQYMGGAIPGIMISTEALFQRAKSLPRVDIIRPSAI 179  
Qy 175 VVGKTEENIRLGVSVALEGIIGRIKEYVGDLP--VLTGGOSKIVKMIKH-EIFDE 232  
Db 180 IARTVNSMQAGIYGVGLVDEIVTRMKAESKDAPRVATGGLASLAPESKTTAEV 239





Db 180 SVVGKNTVECMQAGAVFGAGLVGLVGRVREDVSGFSDHVAIVATGHTAPLLPEL- 238  
QY 227 HEI--FDEDLTIKGV 239  
Db 239 HTVDHYDQHLTLOGL 253

## RESULT 14

US-09-813-453A-10  
; Sequence 10, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Treponema pallidum  
US-09-813-453A-10

Query Match 22.2%; Score 285; DB 9; Length 273;  
Best Local Similarity 32.4%; Pred. No. 3.7e-20;  
Matches 84; Conservative 43; Mismatches 104; Indels 28; Gaps 9;

QY 3 LLDVGNTHSVFTED--GKTFRR--WRLSGVQFQTEDELFSHLHPLGLGDAMREIKGIG 58  
Db 2 LLDVGNTHSVFTED--GKTFRR--WRLSGVQFQTEDELFSHLHPLGLGDAMREIKGIG 58  
QY 59 -----VASVPTQNTVIERFSQKVFHISP-----WVKAKNGCVKWNKPN--SEVGA 104  
Db 57 RASLRDAFISVWPVLTXTIADAVAQISGVQPVVFPWYAEH--LPVRIPEPVRAEIGT 113  
QY 105 DRVANVAFVKEYGKNGIIIDMGATTTVDLV--VNGSYEGGAILPGFFMMVHSLFRGTAKL 163  
Db 114 DLVANAAVYHFRSACVYVPCGTALTFTAVDGTGLIQVATAPGLRTAVQSLHRTGTAOL 173  
QY 164 PLVEVKPADVVGKDEENIRLGVVNGSVYALEGIIGRI--KEVYGDLPVLTGGOSKIVK 222  
Db 174 PLVPLALPDSVLGKDTTHAVQGVVVGRTITFIRAMIAOCQKELGCRCAAVITGGLSRLFS 233  
QY 223 DMKHEIFDEDLTIKGVYH 241  
Db 234 SEVDPPPIDAQLTSLGLAH 252

## RESULT 15

US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01

; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 22.0%; Score 282; DB 9; Length 272;  
Best Local Similarity 31.2%; Pred. No. 7.3e-20;  
Matches 79; Conservative 49; Mismatches 107; Indels 18; Gaps 7;  
QY 3 LLDVGNTHSVFTSI---TEDGKTFRRWRLSTGVQFQTEDELFSHLHPLGLGDAMREIKGIG 58  
Db 3 LAIDVRNTHTVVGLLSGKMEHAKVQVQWRIRTESEVTADELALTIDGLIGEDSERLTGTA 62  
QY 59 VASVPTQNTVIERFSQKVFHISP---IWVKAKNGCVKWNKPNSEVGDARVANVAVFK 115  
Db 63 ALSTVPSVLHEVRIMLDQYWPSPVPHVLIIEFVGRTG--IPLLDVDPKFEVGDARIVNCLAAVD 121  
QY 116 EYKNGIITDMGTATTVDLV--VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFV 174  
Db 122 RFRKAAIVVDFGSSICVDVVSAGFELGGAIAPGVQVSSDAAARSAAALRRVELARPSV 181  
QY 175 VGKDEENIRLGVVNGSVYALEGIIGRIKE-----VYGDLPVLTGGOSKIVKDMIKHE 228  
Db 182 VGKNTVECMQAGAVFGAGLVGLVGRVREDVSGFSDHVAIVATGHTAPLLPEL-HT 240  
QY 229 I--FDEDLTIKGV 239  
Db 241 VDHIDQHLTLOGL 253

Search completed: June 24, 2003, 22:23:56  
Job time : 11.2158 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.97264 Seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITEDG.....HEIFDEDLTIKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1281	100.0	246	2 D72320	conserved hypothetical
2	389.5	30.4	254	2 F83660	hypothetical prote
3	378.5	29.5	259	2 AF1102	conserved hypothetical
4	377.5	29.5	259	2 AF1464	conserved hypothetical
5	369.5	28.8	233	2 S66100	conserved hypothetical
6	360.5	28.1	273	2 E97293	probable transcrip
7	359.5	28.1	261	2 B87489	transcription acti
8	342.5	26.7	265	2 T36391	hypothetical prote
9	289	22.6	274	2 H86937	conserved hypothetical
10	286	22.3	272	2 A70955	hypothetical prote
11	285	22.2	273	2 D71326	conserved hypothetical
12	260	20.3	262	2 E75516	conserved hypothetical
13	204.5	16.0	224	2 A99571	conserved hypothetical
14	203	15.8	262	2 F70165	conserved hypothetical
15	193.5	15.1	229	2 E70465	hypothetical prote
16	179	14.0	223	2 G71887	hypothetical prote
17	165	12.9	223	2 F64627	hypothetical prote
18	152	11.9	267	2 I40327	baf protein - Bord
19	150.5	11.7	592	2 H82031	probable biotin-la
20	146	11.4	592	2 B81009	Bira protein/Bvg a
21	145	11.3	248	2 H83111	hypothetical prote
22	139.5	10.9	209	2 H81382	hypothetical prote
23	138	10.8	257	2 S75559	hypothetical prote
24	124.5	9.7	276	2 AI2292	hypothetical prote
25	103	8.0	242	2 A82637	conserved hypothetical
26	100.5	7.8	313	2 E97221	transcription regu
27	98	7.7	315	2 F70203	xylose operon regu
28	95	7.4	780	2 B84004	ATP-dependent DNA
29	94.5	7.4	287	2 C82158	ROK family protein

ALIGNMENTS

RESULT 1

D72320  
Conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: D72320  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; H  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genom  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: D72320  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-246 <ARN>  
A;Cross-references: GB:AE001754; GB:AE000512; NID:94981417; PIDN:AAD35964.1; PID:94  
A;Experimental source: strain MSB  
C;Genetics:  
A;Gene: TM0883  
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31C

Query Match 100.0%; Score 1281; DB 2; Length 246;  
Best Local Similarity 100.0%; Pred. No. 4.3e-99;  
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLLDGAMREIKGIVA 60  
Db 1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSLHPLLDGAMREIKGIVA 60  
QY 61 SVYPTQNTVIERFSQKVFHISPIWVKARNGCVKNVKNPSEVGADRVANVAFVKEYGKN 120  
Db 61 SVYPTQNTVIERFSQKVFHISPIWVKARNGCVKNVKNPSEVGADRVANVAFVKEYGKN 120  
QY 121 GIIDMGTTATVDLVVNGSYEGGAILFCFFMMVHSLFRGTAKPLVEVKPADVFVVGKDFE 180  
Db 121 GIIDMGTTATVDLVVNGSYEGGAILFCFFMMVHSLFRGTAKPLVEVKPADVFVVGKDFE 180  
QY 181 ENIRLGVVNGSVYALEGIIGRIKEYGDLFVVLGGQSKIVKDMIKHIFEDDLTIKGYV 240  
Db 181 ENIRLGVVNGSVYALEGIIGRIKEYGDLFVVLGGQSKIVKDMIKHIFEDDLTIKGYV 240  
QY 241 HFCFGD 246  
Db 241 HFCFGD 246

RESULT 2

F83660  
Hypothetical protein BH0086 [Imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: F83660  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, R.; Fuji, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: F83660  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <STO>  
A:CROSS-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BA03805.1; GSPDB:GN00177  
A:Experimental source: strain C-125  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.4%; Score 389.5; DB 2; Length 254;  
Best Local Similarity 35.6%; Pred. No. 5.7e-25;  
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;  
QY 1 MYLLVDGNTSHVSFSDTGKTRFRWRLSTGVFQTEDE-----LFSLHPLLGDMRE 53  
1 MILVIDGNTNLVGVQDETUVHWRLLATSRQKTEDEYAMTVRSFLDH-----AGLQFOD 56  
QY 54 IKGIGVASVPTQNTVIERFSQYFHSPIWV--KAKNGCVKWNK--NPSEVGADRVAN 109  
57 IDGIVISSVPPMFSLEQCKKYPHVPIMIGPKTKGL---NIKYDNPKEVGADRVN 113  
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
114 AVAATELYGPAIVVDFGTATTCYIDEGAGVGGIAPGIMISTEALYHRAKLPRIE 173  
QY 169 KPADPVVGGKDEENIRLGVVNGSVYALEGIIIGRIK-EYVGDLPVLTGQSKIV-KDMIK 226  
174 AKPVGVGNTFIDSMQSGIFYGVSDVGVKRMKAQAESEPKVIATGGGLAKLIGTSET 233  
QY 227 HEIFEDLTIKGV 239  
234 IDVIDSFLTKGL 246

RESULT 3  
AF1102  
conserved hypothetical protein lmo0221 [Imported] - *Listeria monocytogenes* (strain EGD-e  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1102  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
.; Jones, L.M.; Karst, U.  
ence 294, 849-852, 2001  
Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1102  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:CROSS-references: GB:NC\_003210; PIDN:CAB00748.1; PID:g16409586; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 29.5%; Score 378.5; DB 2; Length 259;  
Best Local Similarity 35.6%; Pred. No. 4.8e-24;  
Matches 90; Conservative 52; Mismatches 90; Indels 21; Gaps 7;  
QY 1 MYLLVDGNTSHVSFSDTGKTRFRWRLSTGVFQTEDE-----FSLHPLLGDM 51  
1 MILVIDGNTCTGVGYEKQLLKHWRMTDRHRTSDELGMTVNLFFSYANLTP----- 54  
QY 52 REIKGIGVASVPTQNTVIERFSQYFHSPIWV--KAKNGCVKWNKPNSEVGADRVAN 109

Db 55 SDIQGLIISSVPPIMHAMETMCVRYFNIRPLIVGPKITG-LNLKVDNPREIGSDRVN 113  
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
Db 114 AVAASEEYGPVIVDFGTATTCYIDEGAGVGGIAPGIMISTEALYHRAKLPRIE 173  
QY 169 KPADPVVGGKDEENIRLGVVNGSVYALEGIIIGRIK-EYVGDLPVLTGQSKIV-KDMIK 227  
Db 174 AESSQIIGKTSVSSMOAGIFYGVQCEGIIAEMKKOSNAPVAVVATGGGLARMITKSSA 233  
QY 228 -EIFEDLTIKGV 239  
Db 234 VDILDPLFLKGL 246

RESULT 4  
AF1464  
conserved hypothetical protein lln0253 [Imported] - *Listeria innocua* (strain Clp112  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:CROSS-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 29.5%; Score 377.5; DB 2; Length 259;  
Best Local Similarity 35.2%; Pred. No. 5.8e-24;  
Matches 89; Conservative 53; Mismatches 90; Indels 21; Gaps 7;  
QY 1 MYLLVDGNTSHVSFSDTGKTRFRWRLSTGVFQTEDE-----FSLHPLLGDM 51  
Db 1 MILVIDGNTCTGVGYEKQLLKHWRMTDRHRTSDELGMTVNLFFSYANLTP----- 54  
QY 52 REIKGIGVASVPTQNTVIERFSQYFHSPIWV--KAKNGCVKWNKPNSEVGADRVAN 109  
Db 55 SDIQGLIISSVPPIMHAMETMCVRYFNIRPLIVGPKITG-LNLKVDNPREIGSDRVN 113  
QY 110 VVAFVKEYGKNGIIDMGTTATVDLV-VNGSEGGAILPGFFMMVHSLFRGTAKLPLVEV 168  
Db 114 AVAASEEYGPVIVDFGTATTCYIDEGAGVGGIAPGIMISTEALYHRAKLPRIE 173  
QY 169 KPADPVVGGKDEENIRLGVVNGSVYALEGIIIGRIK-EYVGDLPVLTGQSKIV-KDMIK 227  
Db 174 AESSQIIGKTSVSSMOAGIFYGVQCEGIIAEMKKOSNAPVAVVATGGGLARMITKSSA 233  
QY 228 -EIFEDLTIKGV 239  
Db 234 VDILDPLFLKGL 246

RESULT 5  
S66100  
conserved hypothetical protein yacB - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000  
C:Accession: S66100; E69740  
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994

[illegible]







Search completed: June 24, 2003, 22:03:06  
Job time : 9.97264 secs

WJ	400	ZBNH	28	NFCT	PTN	FTG	HTG	AKR	INAY		GDTF	V	LQ	CCQ	KVY	KRM	HNHJF	DEDDJIK	253
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
D6	162	RECVLGG	AYR	RETEFIK	SJKL	WRKVFR	KFKVTW	TGGESKYFS	----	KFGVDPLA	WHRG	218							

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.2563 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHVSFSTEDG.....HEIFDEDLTKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 908470 seqs, 133250620 residues  
-total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1281	100.0	246	AAU91156	Thermotoga maritima
2	405.5	31.7	258	AAU01243	B. subtilis novel
3	405.5	31.7	258	AAU91149	Bacillus subtilis
4	399.5	31.2	262	AAU91170	Pantothenate kinas
5	389.5	30.4	254	AAU91171	Pantothenate kinas
6	378.5	29.5	259	ABB47661	Listeria monocytog
7	375.5	29.3	258	AAU91172	Pantothenate kinas
8	369.5	28.8	233	AAU91163	Pantothenate kinas
9	361.5	28.2	255	AAU91154	Geobacter sulfurre
10	359.5	28.1	260	AAU91173	Pantothenate kinas

11	342.5	26.7	265	23	AAU91151	Streptomyces coeli
12	333.5	26.0	250	23	AAU91150	Clostridium acetob
13	321	23.1	256	23	AAU91175	Pantothenate kinas
14	295.5	23.1	258	23	AAU91153	Rhodobacter capsul
15	286	22.3	272	23	AAU91152	Mycobacterium tube
16	285	22.2	273	23	AAU91157	Treponema pallidum
17	282	22.0	272	22	AAAG81225	Mycobacterium tube
18	279.5	21.8	219	23	AAU91176	Pantothenate kinas
19	260	20.3	262	23	AAU91155	Deinococcus radiop
20	229.5	17.9	241	23	AAU91179	Pantothenate kinas
21	228.5	17.8	257	23	AAU91174	Pantothenate kinas
22	221	17.3	212	23	AAU91177	Pantothenate kinas
23	203	15.8	262	23	AAU91158	Borrelia burgdorfe
24	193.5	15.1	229	23	AAU91159	Aquifex aeolicus p
25	165	12.9	223	19	AAW98422	H. pylori GHPO 344
26	165	12.9	223	23	AAU91161	Helicobacter pylor
27	165	12.9	223	23	AAU91181	Pantothenate kinas
28	162	12.6	244	23	AAU91168	Pantothenate kinas
29	152.5	11.9	455	20	AAV38617	Neisseria gonorrhoe
30	152.5	11.9	455	21	AAV74908	Neisseria gonorrhoe
31	152.5	11.9	460	23	AAU91167	Pantothenate kinas
32	152.5	11.9	592	20	AAV38618	Neisseria gonorrhoe
33	152.5	11.9	592	21	AAV74911	Neisseria gonorrhoe
34	152	11.9	267	23	AAU91162	Bordetella pertussis
35	150.5	11.7	455	21	AAV74910	Neisseria meningit
36	150.5	11.7	592	20	AAV38616	Neisseria meningit
37	150.5	11.7	592	21	AAV74913	Neisseria meningit
38	150.5	11.7	592	23	AAU91166	Pantothenate kinas
39	149.5	11.7	389	21	AAV74909	Neisseria meningit
40	146	11.4	592	20	AAV38615	Neisseria meningit
41	146	11.4	592	21	AAV74912	Neisseria meningit
42	146	11.4	592	23	AAU91169	Pantothenate kinas
43	145.5	11.4	249	23	AAU91182	Pantothenate kinas
44	145	11.3	248	23	AAU91164	Pantothenate kinas
45	139.5	10.9	209	23	AAU91165	Pantothenate kinas

## ALIGNMENTS

## RESULT 1

AAU91156

ID AAU91156 standard; Protein; 246 AA.

XX AAU91156;

AC AAU91156;

DT 05-JUN-2002 (first entry)

XX Thermotoga maritima pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;

XX Pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Thermotoga maritima.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein

XX PS Claim 10; Page 74-75; 128pp; English.

XX CC The invention describes assays for identifying a (potential) antibiotic

CC comprising contacting an assay composition comprising a pantothenate

CC kinase (Coax) protein with a test compound, and determining the ability

CC of the test compound to inhibit the activity of the Coax protein, an

CC essential enzyme for the production of coenzyme A. Coax protein is a

CC valuable target for identifying bactericidal compounds. Coax modulating

CC agents can be used in an infectious animal model to determine the

CC efficacy, toxicity, or side effects of treatment with such an agent. This

CC is the amino acid sequence of a pantothenate kinase (Coax) protein

CC described in the invention.

XX SQ Sequence 246 AA;

Query Match 100.0%; Score 1281; DB 23; Length 246;

Best Local Similarity 100.0%; Pred. No. 7.9e-133; Indels 0; Gaps 0;

Matches 246; Conservative 0; Mismatches 0;

1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMREIKIGVA 60

1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMREIKIGVA 60

61 SVVPTQNTVIERFSOKYFPHISPIWVAKNGCVKNVKNPSEVGADRVANVAVFVKEYGKN 120

61 SVVPTQNTVIERFSOKYFPHISPIWVAKNGCVKNVKNPSEVGADRVANVAVFVKEYGKN 120

121 GIIDMGATTATVDLVVNGSYEGGAILPGFFMVMHSLFRGTAKLPLVEVKPADVVGKDT 180

121 GIIDMGATTATVDLVVNGSYEGGAILPGFFMVMHSLFRGTAKLPLVEVKPADVVGKDT 180

181 ENIRLGVNGSVYALEGIIGRIKEYVDLPVLTGGQSKIVKDMIKHEIFDEDLTIKGVY 240

181 ENIRLGVNGSVYALEGIIGRIKEYVDLPVLTGGQSKIVKDMIKHEIFDEDLTIKGVY 240

241 HECFGD 246

241 HECFGD 246

RESULT 2

AAU01243

ID AAU01243 standard; Protein; 258 AA.

XX AC AAU01243;

XX DT 18-JUL-2001 (first entry)

XX B. subtilis novel pantothenate kinase encoded by the gene coax.

KW Pantothenate kinase; coax; pantothenate biosynthesis; vitamin B5;

KW nutritional supplement; panto-compound; pantoate.

OS Bacillus subtilis.

XX WO200121772-A2.

XX 29-MAR-2001.

XX 21-SEP-2000; 2000WO-US25993.

XX 21-SEP-1999; 99US-0400494.

PR 07-JUN-2000; 2000US-0210072.

PR 28-JUL-2000; 2000US-0221836.

PR 24-AUG-2000; 2000US-0227860.

XX (OMNI-) OMNIGENE BIOPRODUCTS.

XX Yocum RR, Patterson TA, Hermann T, Pero JG;

XX WPI: 2001-218644/22.

DR N-PSDB; AA500984.

XX PT New recombinant microorganism which overexpress a Bacillus subtilis

PT pantothenate biosynthetic enzyme, useful for the high yield production

PT of panto-compounds such as pantothenate and pantoate -

XX Example 14; Fig 23; 292pp; English.

PS The sequence represents a novel B. subtilis pantothenate kinase (encoded

CC by gene coax), an enzyme of the pantothenate biosynthetic pathway.

CC Pantothenate, also known as vitamin B5, is used as a nutritional

CC supplement in mammals and humans. The invention concerns methods of

CC producing recombinant microorganisms overexpressing at least one Bacillus

CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods

CC of producing them are useful for producing a panto-compound such as

CC pantothenate or pantoate, which is a nutritional requirement for

CC of livestock and humans. The methods are also useful for the identification

CC of pantothenate kinase modulators. Panto-compounds are produced at a

CC significantly higher yield than prior art methods and can be produced

CC independent of the need to feed precursors which decreases expense.

XX SQ Sequence 258 AA;

Query Match 31.7%; Score 405.5; DB 22; Length 258;

Best Local Similarity 39.2%; Pred. No. 4e-36; Indels 25; Gaps 9;

Matches 100; Conservative 42; Mismatches 88;

1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMRE 53

1 LLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDELFSLHPLLLGDAMRE 56

54 IKGIGVASVPTQNTVIERFSOKYFPHISP--IWKAKNGCVKNVKN--NPSEVGADRVAN 109

57 IDGIILSSVVPPTMFALERMCTKTFHIEPQIVGPMKTGL---NIKVDNPKVEGADRVN 113

110 VVAFVKEYGKNGIIDMGATTATVDLV-VNGSYEGGAILPGFFMVMHSLFRGTAKLPLVEV 168

114 AVAAIHLGNPLIVDFGTATTTCYIDENKQYGGALAPGITITSTEALYSRAKLPRIEI 173

169 KPADFVVGKDTENIRLGVNGSVYALEGIIGRIK-EVYGDLPVLTGGQSKIV---KDM 224

174 TRPDNIIGKNTVSAMQSGILFVGVQVEGIVKRMKQKQDLKVIATGGLAPLIANESDC 233

225 IKHEIFEDEDLTIKGV 239

234 I--DIYDPLFLTKGL 246

RESULT 3

AAU91149

ID AAU91149 standard; Protein; 258 AA.

XX AC AAU91149;

XX DT 05-JUN-2002 (first entry)

XX Bacillus subtilis pantothenate kinase Coax.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;

KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX Bacillus subtilis.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX





```

XX 05-JUN-2002 (first entry)
XX Pantothenate kinase (Coax) #10.
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Bacillus stearothermophilus.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum RR, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54193.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein.
XX Claim 10; Page 101-102; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 258 AA;
XX
XX Query Match 29.3%; Score 375.5; DB 23; Length 258;
XX Best Local Similarity 37.5%; Pred. No. 8.2e-33;
XX Matches 94; Conservative 48; Mismatches 92; Indels 17; Gaps 9;
XX
XX 1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVQTEDELFSLHPL--GDAMREIKGI 57
XX 1 MIFVLVDGNTVTLGVYDGLKHHWRIETSRKTEDEYGMKALLNHVGLQFSDIRGI 60
XX
XX 58 GVASVVPQTONTVIERFSQKFFHISPIW--KAKNGC-VKWNKPNSEVGADRVANVAFV 114
XX 61 IISVVVPPINFALERCLYFKPLVIGPGIKTGDIKYD--NPREVGGADRVNNAVAGI 118
XX 115 KEYKNGIILDMGTATTVDLV-VNGSYEGGAILPGFFMVMHSLFRGTAKLPLVEVKPADF 173
XX 119 HYGSPLLIIVDGTATTCYINHEKQYMGGAIPAGIMISTEALFARAALKPRIETARPDD 178
XX 174 VVGKDTPEENIRLGVNNGSVYALEGGIIRKEVYGDLP--VVLFGGQSKIV--KDMIKHE 228
XX 179 IIGKNTVSAQAGILYGVQVEGIVSRMK-AKSKIPPVKPIATGGLAPLASESDII--D 235
XX 229 IFEDETITKGV 239
XX 236 VVDPFLTLAQL 246
XX
XX RESULT 8
XX AAU91163
XX ID AAU91163 standard; Protein; 233 AA.

```

```

XX AAU91163;
XX 05-JUN-2002 (first entry)
XX Pantothenate kinase (Coax) #1.
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX Bacillus subtilis.
XX WO200216601-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US26531.
XX 24-AUG-2000; 2000US-227860P.
XX 20-MAR-2001; 2001US-0813453.
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX Yocum ER, Patterson TA;
XX WPI; 2002-269358/31.
XX N-PSDB; ABK54169.
XX Identifying potential antibiotic or antimicrobial agent, comprises
XX contacting composition comprising pantothenate kinase (Coax) protein
XX with test compound and identifying inhibitor of the Coax protein.
XX Disclosure; Page 81-82; 128pp; English.
XX The invention describes assays for identifying a (potential) antibiotic
XX comprising contacting an assay composition comprising a pantothenate
XX kinase (Coax) protein with a test compound, and determining the ability
XX of the test compound to inhibit the activity of the Coax protein, an
XX essential enzyme for the production of coenzyme A. Coax protein is a
XX valuable target for identifying bactericidal compounds. Coax modulating
XX agents can be used in an infectious animal model to determine the
XX efficacy, toxicity, or side effects of treatment with such an agent. This
XX is the amino acid sequence of a pantothenate kinase (Coax) protein
XX described in the invention.
XX Sequence 233 AA;
XX
XX Query Match 28.8%; Score 369.5; DB 23; Length 233;
XX Best Local Similarity 40.5%; Pred. No. 3.2e-32;
XX Matches 87; Conservative 33; Mismatches 76; Indels 19; Gaps 6;
XX
XX 1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVQTEDE-----LFSLHPLLDGAMRE 53
XX 1 MLLVLVDGNTVTLGVYHDGKLEYHWRIETSRKTEDEFGMLIRSLFDH----SGLMFEQ 56
XX
XX 54 IKGIVASVPTQNTVIERFSQKFFHISPP--IWKAKNGCVKWNK--NPSEVGADRVAN 109
XX 57 IDGIIISVVPPINFALERCLYFKPLVIGPGIKTGDIKYD--NPKYDNPKEVGADRVN 113
XX 110 VVAVKVEGKNGIILDMGTATTVDLV-VNGSYEGGAILPGFFMVMHSLFRGTAKLPLVEV 168
XX 114 AAVAHILYGNPLIIVDGTATTCYIDENKQYMGGAIPAGITISTEALYSRAALKPRIEI 173
XX 169 KPADFVVGKDTPEENIRLGVNNGSVYALEGGIIRIK 203
XX 174 TRPDNIIGKNTVSAQAGILYGVQVEGIVSRMK 208
XX
XX RESULT 9
XX AAU91154
XX ID AAU91154 standard; Protein; 255 AA.
XX AAU91154;

```

XX 05-JUN-2002 (first entry)  
 XX Geobacter sulfurreducens pantothenate kinase Coax.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 KW Geobacter sulfurreducens.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 PA Yocum RR, Patterson TA;  
 WPI; 2002-269358/31.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein.  
 PS Claim 10; Page 72-73; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 255 AA;  
 SQ Query Match 28.2%; Score 361.5; DB 23; Length 255;  
 Best Local Similarity 36.8%; Pred. No. 2.8e-31;  
 Matches 91; Conservative 49; Mismatches 98; Indels 9; Gaps 6;  
 1 MYLLVDVGNTHSVFSTEDGKTRRWRRLSTGVFQTEDE---LPSHLHLPLGDMREIKGI 57  
 1 MLLVIDVGNTHSVFSTEDGKTRRWRRLSTGVFQTEDE---LPSHLHLPLGDMREIKGI 60  
 58 GVASVPTQNTVIERESQKYFHISPIWV--KAKNGCVKNVKNPSEVGDVAVNVAFK 115  
 61 IISVVVPTGTGLERLUSLGFGRPLVGVGKGTG-MPIQYDNPREVGADRIYNVAGYE 119  
 116 EYKNGIILDMGTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFV 174  
 120 KYRTSLIIVDFGTATFDVYNRKGEGCGAIAPLVISTEALFORASKLPRVDIIIPSAI 179  
 175 VKQDTEENRLGVNGSVYALEGIGRIKEVYGDLP-VVLTGQSKIVKMDIKH-BIFDE 232  
 180 IARNTVNSMQAGIYGYGVGLVDEIVTRMKAESKADPRVIATGCLASLIAPESKTIAVEE 239  
 233 DLTIKGV 239  
 240 YLTLEGL 246  
 XX RESULT 10  
 ID AU91173  
 XX AU91173 standard; Protein; 260 AA.

AC AU911173;  
 XX 05-JUN-2002 (first entry)  
 XX Pantothenate kinase (Coax) #11.  
 DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 XX pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 KW Geobacter crescentus.  
 XX WO200216601-A2.  
 XX 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 PA Yocum RR, Patterson TA;  
 WPI; 2002-269358/31.  
 XX N-PSDB; ABK54194.  
 XX Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein.  
 PS Claim 10; Page 102-103; 128pp; English.  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 260 AA;  
 SQ Query Match 28.1%; Score 359.5; DB 23; Length 260;  
 Best Local Similarity 35.5%; Pred. No. 4.8e-31;  
 Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;  
 1 MYLLVDVGNTHSVFSTEDGKTRRWRRLSTGVFQTEDELFSLHPLLL---GDAMREIKGI 57  
 1 MLLAIEQGNNTWFAIHGASVWQWRSATSTRTADEYVWVLSQMLGFGRAIDAV 60  
 58 GVASVPTQNTVIERESQKYFHISPIWV--KAKNGCVKNVKNPSEVGDVAVNVAFK 115  
 61 IISVVVPTGTGLERLUSLGFGRPLVGVGKGTG-MPIQYDNPREVGADRIYNVAGYE 119  
 116 EYKNGIILDMGTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEVKPADFV 172  
 120 VYGPPLVVIDSGTATFDVIAADGAFEGGIAPGINLSMQALHEAAAKLPRIATQRPAGN 179  
 173 FVVGKTEENIRLVNGSVYALEGIGRIKEVYGD-LPVVLTGQSKIVK---DMIKHE 228  
 180 RIVGTDVTSAMQSGVFWGYISLIEGLVARIKAEKGEPTVIATGCVASLFGATDSIDH- 238  
 229 IFDEDLTIKGV 239  
 239 -FDSDLTIRGL 248  
 XX RESULT 11  
 ID AU911151



AAU91151 standard; Protein; 265 AA.  
AAU91151;  
05-JUN-2002 (first entry)  
Streptomyces coelicolor pantothenate kinase Coax.  
Pantothenate kinase; Coax; antibiotic; antimicrobial;  
pantothenate kinase modulator; coenzyme A; bactericidal compound.  
Streptomyces coelicolor.  
WO200216601-A2.  
28-FEB-2002.  
24-AUG-2001; 2001WO-US26531.  
24-AUG-2000; 2000US-227860P.  
20-MAR-2001; 2001US-0813453.  
(OMNI-) OMNIGENE BIOPRODUCTS INC.  
Yocum RR, Patterson TA;  
WPI; 2002-269358/31.  
Identifying potential antibiotic or antimicrobial agent, comprises  
contacting composition comprising pantothenate kinase (Coax) protein  
with test compound and identifying inhibitor of the Coax protein -  
Claim 10; Page 69-70; 128pp; English.  
The invention describes assays for identifying a (potential) antibiotic  
comprising contacting an assay composition comprising a pantothenate  
kinase (Coax) protein with a test compound, and determining the ability  
of the test compound to inhibit the activity of the Coax protein, an  
essential enzyme for the production of coenzyme A. Coax protein is a  
valuable target for identifying bactericidal compounds. Coax modulating  
agents can be used in an infectious animal model to determine the  
efficacy, toxicity, or side effects of treatment with such an agent. This  
is the amino acid sequence of a pantothenate kinase (Coax) protein  
described in the invention.  
Sequence 265 AA;  
Query Match 26.7%; Score 342.5; DB 23; Length 265;  
Best Local Similarity 35.7%; Pred. No. 3.7e-29;  
Matches 92; Conservative 45; Mismatches 98; Indels 23; Gaps 9;  
QY 1 MLLVDVGNTHSVFSITEDGKTRFRRLSTGVFQTEDEL-----FSLHPLLGDMARE- 53  
DB 1 MLLIDVGNTHVTLGLFDGDIVVHWRISTDSRRTADELAVLLQGLMGHPLLGDELGDG 60  
QY 54 IKGIGVASVPTQNTVIERFSQKYPHISPIWKAAGCKVKNV-----KNPSEVGADRVAN 109  
DB 61 IDGTAICATVPSVLHREVRTRYGDVPA-VLVEPG-VKTGVPILTDHPKEVGADRIIN 118  
QY 110 VVAFKEGKNGIIDMGATTVDLV-VNGSVEGGAILPGFMMVHSLFRGTAKLPLEV 168  
DB 119 AVAAVELYGGPAIVVDGTTATFDVARSARGEYIGVIAPIEISVEALGVKGAQRKIEV 178  
QY 169 KPADPVGKGTDEENIRLGVNGSVYALEGIIGRIKEVYGDLP-----VVLTGGSQKIVKDM 224  
DB 179 APRSVIGKNVVEAMQSIGIVGFAGQVDGVVNRMARLADDDVDVTVIATGGLAPMV--L 236  
QY 225 IKHEIFDEE---LTIKV 239  
DB 237 GESSVIDEHEPWLTLML 254

AAU91150  
ID AU91150 standard; Protein; 250 AA.  
XX  
AC AAU91150;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Clostridium acetobutylicum pantothenate kinase Coax.  
XX  
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX  
OS Clostridium acetobutylicum.  
XX  
PN WO200216601-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 24-AUG-2001; 2001WO-US26531.  
XX  
PR 24-AUG-2000; 2000US-227860P.  
PR 20-MAR-2001; 2001US-0813453.  
XX  
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX  
PI Yocum RR, Patterson TA;  
XX  
DR WPI; 2002-269358/31.  
XX  
PT Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
XX  
PS Claim 10; Page 68-69; 128pp; English.  
XX  
SQ The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.  
XX  
SQ Sequence 250 AA;  
Query Match 26.0%; Score 333.5; DB 23; Length 250;  
Best Local Similarity 39.1%; Pred. No. 3.4e-28;  
Matches 95; Conservative 39; Mismatches 72; Indels 37; Gaps 11;  
QY 3 LLVDVGNTHSVFSITEDGKTRFRRLSTGVFQTEDE-----LFSH--LHPLLGDMARE 53  
DB 20 LVLDVGNTHVILGIYNDTKLTAERLSTDLVRSADVEYGIQVMNLFQODKLDPTL----- 73  
QY 54 IKGIGVASVPTQNTVIERFSQKYPHISPIW--KAKNGCVKNV--NPSEVGADRVAN 109  
DB 74 VEGVISSVVPNIMYSLEHMIRKYFKINPLVWVGPKTGT---NIRDNRKVEGADRIIN 130  
QY 110 VVAFKEGKNGIIDMGATTVDLV-VNGSVEGGAILPGFMMVHSLFRGTAKLPLEV- 167  
DB 131 AVAAHEIYKRSLIIDFGTATTCVAVRENGDYLGGAICPGIKVSSEALFEKAKLPVEL 190  
QY 168 VKPADPVGKGTDEENIRLGVNGSVYALEGIIGRIKEVYGDLP-----VVLTGGSQ 218  
DB 191 IKPA-YAICKNTISSIQSIGIVYLRQVKYLFELKE---NLPDGRTRTSLVLTATGGLA 246  
QY 219 KIV 221  
DB 247 KLI 249



D<sup>b</sup> 239 HTVDHYDQHLTLQGL 253  
| : :|: ||::|:

Search completed: June 24, 2003, 21:46:17  
Job time : 26.2563 secs

D**b** 232 DLFDKVEDDLTMHGL 246

## RESULT 15

AAU911152  
ID AAU911152 standard: protein: 272 AA.

AAU91152:

AA	DT	05-JUN-2002 (first entry)
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

DE *Mycobacterium tuberculosis* pantothenate kinase Coax.

XX  
KW  
Pantothenate kinase; CoaX; antibiotic; antimicrobial;

XX  
KW  
XX  
pallidocnemate kinase modulator; coenzyme A; bacterial compound.

XX  
03 MY COMPACTETUM CABINETCLOSURE:

WO200216601-A2.

28-FEB-2002.

24-AUG-2001: 2001WO-US26531.

XX PR 24-AUG-2000: 2000US-227860P.

PR 20-MAR-2001; 200103-0813433:  
XX

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX  
PI  
Yocum RR. Patterson TA:

XX  
DR WPT: 2002-269358/31.

AA Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -  
PT

PS Claim 10; Page 70-71; 128pp; English.

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence	272 AA:
S0	

Query Match 22.3%; Score 286; DB 23; Length 272;

Best Local Similarity	ST-48, Recd. No. 0.70 257	Indels	Gaps
Matches 80; Conservative	49; Mismatches 108;	18;	7;

QV 1 MYLLVDVGNTHSVFSI-----TEDGKTFRRWRRLSTGVFQTEDELFSHLHPLLGDAMREIKG 56

1 ML LAIDVRNTHTVVGLLSGMEKHEAKVVOQWRIRTESEVTADELALTIDGLIGEDSERLTG 60

57 IGVASVWPTONTVIERFSOKYFHSIP - - - IWVKAKNGCVKWNVKNPSEVGADRVANVAF 113

Db 61 TAALSTVPVLHEVRIMLDQYWPSPVPHVLTIEPGVRTG-IPLLVDPNPKEVGADRIVNC<sub>LAA</sub> 119

114 VKEYGKNGI IIDMGATTVDLV-VNGSYEGGAILPGFFMMVHSLERGTAKLPLVEVKPAD 172

Db 120 YDRERKAAIWDGSSICVDVWSAKGEFTGGATAPGVVOVSSDAAARSAAALRRVELARPR 179

173 FWCKDTEFNRI GWNGSWALEGIIGRIKE-----VYGDLPWVLTGGOSKTVKDMTK 226

db 180 SWVCKNTTVECMOAGAVEGFAGLVDGLVGRITREDVSGFSVDHDAIVATGHTAP[L.I.PEL]- 238

227 HEI--EDEDI.TIKCV 239

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.27862 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSITDGD.....HEIFDEDLIKGVYHCFDGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 112892 seqs, 41476328 residues 112892  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description
1	369.5	28.8	1 YACB_BACSU	P37564 bacillus su
2	152	11.9	2 BAF_BORPE	Q45338 bacillus bor
3	94	7.3	3 AAT_STRVG	Q60013 streptococ
4	94	7.3	4 RPOD_ANASP	P22705 anabaena sp
5	93.5	7.3	5 RPOD_NOSCO	P14564 nostoc comm
6	92	7.2	6 Y840_METJA	Q58250 methanococc
7	91.5	7.1	7 AROA_METJA	Q57925 methanococc
8	90.5	7.1	8 RPOC_LEUME	P94892 leuconostoc
9	89.5	7.0	9 RPOC_LEUPS	P94899 leuconostoc
10	89	6.9	10 NDAM_ACACA	Q37373 acanthamoeb
11	88.5	6.9	11 LE11_ARCFU	Q29305 archaeglob
12	88	6.9	12 RPOC_BACHD	Q67943 aquifex ae
13	88	6.9	13 RPOC_BACHD	Q929m1 bacillus ha
14	87	6.8	14 RPOC_OENOE	P95405 oenococcus
15	86	6.7	15 CTRP_DROME	Q46036 drosophila
16	86	6.7	16 FLID_BACHD	P95816 streptococc
17	85.5	6.7	17 HSCA_BUCAP	Q9K6W0 bacillus ha
18	85.5	6.7	18 RPOC_BUCAP	O51883 buchnera ap
19	85.5	6.7	19 RPOC_LISMO	Q89a96 listeria mo
20	85.5	6.7	20 RPOC_FORGI	P42375 porphyromon
21	85	6.6	21 RPOC_LISIN	P77879 listeria in
22	84.5	6.6	22 TRGS_ECOLI	Q00184 escherichia
23	84	6.6	23 RPOC_WEIHE	Q66177 weissella h
24	84	6.6	24 RPOC_METJA	Q58142 methanococc
25	83.5	6.5	25 PSAB_PROMA	Q9RdV0 pyrochlocco
26	83.5	6.5	26 DPOL_PYRAB	P77916 pyrococcus
27	83.5	6.5	27 CA36_CHICK	P15989 gallus gall
28	83.5	6.5	28 HXK1_HUMAN	P19367 homo sapien
29	83	6.5	29 HGDC_ACIFE	P11568 acidaminoco
30	82.5	6.4	30 PRIM_CLOAB	P33655 clostridium
31	82.5	6.4	31 YED0_YEAST	P39987 saccharomyc
32	82.5	6.4	32 LCF3_YEAST	P39002 saccharomyc
33	82.5	6.4		

34	82.5	6.4	1047	1 CARB_THEAC	Q9hk17 thermoplas
35	82	6.4	437	1 SUCB_NEOFR	P33587 neocallinas
36	82	6.4	541	1 DNK6_MYCHY	O34191 anaplasma p
37	82	6.4	600	1 HSCA_BUCAI	O49539 mycoplasma
38	82	6.4	611	1 G3PC_PETCR	P57660 buchera ap
39	81.5	6.4	336	1 G3PC_PETCR	P26519 petroselinu
40	81.5	6.4	519	1 GALT_THETN	Q8r8r6 thermoanae
41	81	6.3	1052	1 RPOC_BACAN	P77819 bacillus an
42	80.5	6.3	335	1 G3P_CHLMU	Q9pjp6 chlamydia m
43	80.5	6.3	517	1 LEUL_NEJMA	O9juk6 neisseria m
44	80.5	6.3	591	1 CO8B_HUMAN	P07358 homo sapien
45	80.5	6.3	1055	1 RPOC_PEDAC	P77917 pediococcus

ALIGNMENTS

RESULT 1  
ID YACB\_BACSU STANDARD; PRG: 233 AA.  
AC P37564;1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; Pubmed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; Pubmed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningstein G., Krogh J., Lazarevic V.,  
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D26185; BAA05305.1; -  
DR EMBL; Z99104; CAB11846.1; -  
DR Subtilist; BG10133; YacB.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRfams; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;  
  
Query Match 28.8%; Score 369.5; DB 1; Length 233;  
Best Local Similarity 40.5%; Pred. No. 5.1e-24;  
Matches 87; Conservative 33; Mismatches 76; Indels 19; Gaps 6;  
  
QY 1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDE-----LFSHLHPLLDGAMRE 53  
Db 1 MLLVIDVGNTHSVFSTEDGKTFRRWRLSTGVFQTEDEFGMLRLSLFDH-----SGLMFEQ 56  
  
QY 54 IKGIGVASVVPQTNTVIERFSQKYPHSP--IWKAKNGCVKWNK--NPSEVGADRVAN 109  
Db 57 IDGIITSSVVPPIFALERMCTKYFHEPQIVGPMKTLG--NIKVDNPKVEGADRIYN 113  
  
QY 110 VYAFVKEYGKNIIDMGATTVDLV-VNGSEGGAILGFFFMVHSLFRGTAFLPLEV 168  
Db 114 AVAAHLHNGNPLVVDFTGTCYIDENKQVNGGAIAPGINTSTALYSRAAKLPRIE 173  
  
QY 169 KPADFVVGKDTENIRLGVNGSVYALEGIIGRIK 203  
Db 174 TRPDNICKNTVSAMQSGILFGVVGQVEGIVRMK 208  
  
RESULT 2  
ID BAF\_BORPE STANDARD; PRT; 267 AA.  
AC Q45338; Q45373;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
GN Baf.  
Bordetella pertussis.  
Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
Bordetella.  
NCBI\_TaxID=520;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-BP504;  
RX MEDLINE=9325323; PubMed=7601846;  
RA Desnazer D., Wood G.E., Friedman R.L.;  
RT "Identification of a Bordetella pertussis regulatory factor required for transcription of the pertussis toxin operon in Escherichia coli."  
RL J. Bacteriol. 177:3801-3807(1995).  
RN [2]  
RP SEQUENCE OF 1-38 FROM N.A.  
RC STRAIN-BP504;  
RA Wood G.E., Friedman R.L.;  
RT "Identification of a bira homolog in Bordetella pertussis."  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE OF 239-267 FROM N.A.  
RC STRAIN-BP536;  
RX MEDLINE=96419162; PubMed=8821935;  
RA Allen A.G., Maskell D.J.;  
RT "The identification, cloning and mutagenesis of a genetic locus

required for lipopolysaccharide biosynthesis in Bordetella pertussis.";  
Mol. Microbiol. 19:37-52(1996).  
-!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF RNA POLYMERASE.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U12020; AAA75361.1; -  
DR EMBL; AF016461; AAC68834.1; -  
DR EMBL; X90711; CAA62242.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Transcription regulation; Activator.  
SQ SEQUENCE 267 AA; 27845 MW; 336A615F6B57901 CRC64;  
  
Query Match 11.9%; Score 152; DB 1; Length 267;  
Best Local Similarity 25.0%; Pred. No. 1.1e-05;  
Matches 62; Conservative 36; Mismatches 88; Indels 62; Gaps 10;  
  
QY 1 MYLLVDVGNTHSVFSTEDGKTFRRWRLSTGVFQTE-----DELFSH 42  
Db 1 MLLIDSGNS-----RLKVGWDFDPAPQAREPAPVAFDNLDDALGRW 44  
  
QY 43 LHPLLDGAMREI-----KGIGVASVVPQTNTVIERFSQKYPHSP--IWKAKNGCVK 94  
Db 45 LATLPRRQALGVNAGLARGAIAATLRAGCDI-----RWLRAQPLAMGLNG----- 95  
  
QY 95 NVKNPSEVGADR--VANVAVFVKEYGKNGIIDMGATTVDLV-VNGSEGGAILPGFF 150  
Db 96 -YRNPDLGADRWACWVGVLARQPSVHPPLLVASFATTTDITGPNVFPGLILPGPA 154  
  
QY 151 MMVHSLFRGTAFLPLEVVPADPVGVGKDTENIRLGVNGSVYAL--EGIIIGRIKEYVD 208  
Db 155 MMGALAYGTAHLPLADGLVADYPI--DTHQATASGIAAQAAGVIRQWLAGR--QRYGQ 210  
  
QY 209 LPFVLTGG 216  
Db 211 APEIYVAG 218  
  
RESULT 3  
AAT\_STRVG STANDARD; PRT; 397 AA.  
ID AAT\_STRVG  
AC Q60013;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Aspartate aminotransferase (EC 2.6.1.1) (Transaminase A) (ASPART).  
GN ASPC OR AAT.  
OS Streptomyces virginiae.  
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1961;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96257210; PubMed=8675024;  
RA Katayama M., Sakai Y., Okamoto S., Ihara F., Nihira T., Yamada Y.;  
RT "Gene organization in the ada-rpL region of Streptomyces virginiae.";  
RL Gene 171:135-136(1996).  
CC -!- CATALYTIC ACTIVITY: L-aspartate + 2-oxoglutarate -> oxaloacetate + L-glutamate.  
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
-----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D50624; BAA09299.1; -  
CC HSP; O56232; LBW;  
CC InterPro: IPR001176; ACC\_synthase.  
CC InterPro: IPR004839; AminoTransf1/2.  
CC InterPro: IPR004838; NHtransf1.  
CC Pfam: PF00155; aminotran\_1.2; 1.  
CC PRINTS: PR00753; ACCSYNTHASE.  
CC PROSITE: PS00105; AA\_TRANSFER\_CLASS.1; 1.  
CC Transferrase; AminoTransferase; Pyridoxal phosphate.  
KW BINDING 236 236 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
FT SEQUENCE 397 AA; 42381 MW; EEFDECEB7D923065 CRC64;

Query Match 7.3%; Score 94; DB 1; Length 397;  
Best Local Similarity 24.5%; Pred. No. 1.3;  
Matches 60; Conservative 38; Mismatches 101; Indels 46; Gaps 15;  
QY 3 LLDVGN-THSVFSTEDGKTRWRRLSTGVFTDELFSHL-----HPLIGDAMR 52  
Db 167 LFWSPNPTGVS-EADAKAIGEMAHEGLVLTDEIYHLVYGEAKFTSLPLVPALR 225  
QY 53 EIKGIGVASVPTQNTVIERFSQYFHSPIWKAAGCVK--WNKPNSEVGDARVAV 110  
Db 226 D-KCIIIV-----NGVAKTYAMTGRVG--WVIAPODVIAKATNLSQSHATSNVNAQV 275  
QY 111 VAPVKEYGKNGIIMGTA-----TTVDLV--VNGSY-----EGG-ALLPGFFMMVHSLF 157  
Db 276 ALAAVSGNLDVAEMKAFDRRQRTQVWMLNEIDGVCTPESGAFYVPSVKELGKEI 335  
QY 158 RGTAKPLVEKVPADVFVGKDTENIRLGVNSVVALEGIIGRIKEVYGDLPVLTGGQ 217  
Db 336 RG--KRQSSVELAAILL-----DEVEVAVVPEAFGTGPGYL-RLSYALGDDELIV--EGV 385  
QY 218 SKIVK 222  
Db 386 SRIQK 390

RESULT 4  
D RPOD\_ANASP STANDARD; PRT; 1355 AA.  
AC P22705;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA-directed RNA polymerase delta chain (EC 2.7.7.6).  
GN RPOC2 OR ALR1596.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriquchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RL cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
RN [2]  
RP SEQUENCE OF 1-63 FROM N.A.  
RX MEDLINE=91258327; PubMed=1904436;  
RA Bergsland K.J., Haselkorn R.;  
RT "Evolutionary relationships among eubacteria, cyanobacteria, and

RT chloroplasts: evidence from the rpoC1 gene of Anabaena sp. strain PCC  
RT 7120.";  
RL J. Bacteriol. 173:3446-3455(1991).  
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
CC {RNA}(N).  
CC -!- SUBUNIT: IN CYANOBACTERIA THE RNA POLYMERASE IS COMPOSED OF FOUR  
CC SUBUNITS: ALPHA, BETA, GAMMA, AND DELTA.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP003586; BAB77962.1; ALT\_INIT.  
CC EMBL; M60831; AAA22034.1; -  
CC PIR; C42361; C42361.  
CC PIR; S14551; S14551.  
CC InterPro: IPR000722; RNA\_pol\_A.  
CC Pfam: PF00623; RNA\_pol\_A.1.  
CC Transferrase; Transcription; DNA-directed RNA polymerase;  
KW Complete proteome.  
SQ SEQUENCE 1355 AA; 147582 MW; 761EE9D92870C961 CRC64;

Query Match 7.3%; Score 94; DB 1; Length 1355;  
Best Local Similarity 22.4%; Pred. No. 5.7;  
Matches 60; Conservative 36; Mismatches 90; Indels 82; Gaps 14;  
QY 5 VDVGNTHS--VFSITEDGKTRWRRLSTGVF-----QTEDELFSHLPLIGDAM 51  
Db 217 IDCCTTGRGIPVRPMTGSKTL--IKLSTRLGRVVGEDVIHPKTEVAPRNTPISSDLA 274  
QY 52 REIKGIGVASVPTQNTVIERFSQYFHSPIWKAAGCVK--WNKPNSEVGDARV 107  
Db 275 KEIKAGVAVVVV-----SPLTCEARSVCQHCYGSALAHAKMVDLGEA 319  
QY 108 ANVAVFKEYGKNGIIMGTA--VDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLP--- 164  
Db 320 VGIIA-AQSIGEPGTLTMTFT-----GGVFTGEVAQQVRSKMDGTIKLPRKL 368  
QY 165 -----LVEVKPADVFVGKDTENIRLGVNSVVALEGIIGRIK 203  
Db 369 RTRTRHTRHGEDALFVESNGIMLEPRK--EGSETPAQEIHTVGTGTTIIVDG--QQVK 424  
QY 204 EYVGD-LPVLTGGOS-----KIVKDM 224  
Db 425 K--GOLLAELVAGGRTTRTNTTEKAVKDV 450

RESULT 5  
D RPOD\_NOSCO STANDARD; PRT; 717 AA.  
AC P14564;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-AUG-1991 (Rel. 19, Last annotation update)  
DE DNA-directed RNA polymerase delta chain (EC 2.7.7.6) (Fragment).  
GN RPOC2.  
OS Nostoc commune.  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=1178;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89197764; PubMed=2495268;  
RA Xie W.-O., Jaeger K., Potts M.;  
RT "Cyanobacterial RNA polymerase genes rpoC1 and rpoC2 correspond to  
RT rpoC of Escherichia coli.";

RL J. Bacteriol. 171:1967-1973(1989).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: IN CYANOBACTERIA THE RNA POLYMERASE IS COMPOSED OF FOUR  
 CC SUBUNITS: ALPHA, BETA, GAMMA, AND DELTA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M29747; AAA35518.1; -;  
 CC PIR; B32838; B32838.  
 CC HSP; Q9KW06; LHOM.  
 CC InterPro; IPR000722; RNA\_pol\_A.  
 CC InterPro; IPR002879; RNA\_pol\_A2.  
 CC Pfam; PF00623; RNA\_pol\_A; 1.  
 CC Pfam; PF01854; RNA\_pol\_A2; 1.  
 CC Transferrase; Transcription; DNA-directed RNA polymerase.  
 CC NON\_TER 717  
 CC SEQUENCE 717 AA; 78244 MW; E5505A9FA37D3C91 CRC64;  
 CC -----  
 CC Query Match 7.3%; Score 93.5; DB 1; Length 717;  
 CC Best Local Similarity 22.2%; Pred. No. 2.9;  
 CC Matches 61; Conservative 35; Mismatches 94; Indels 85; Gaps 13;  
 CC -----  
 CC 4 LVDVGNTH-----SVFSITEDGKTF--RRWRLSTGVF-----QTEDELFSLH 43  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 205 LVDVSVYYSFGDCGPELSIRPTEGAKTLIPLATRLMGRVIGEDVLHPVTKEVIAARN 264  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 44 HPLIGDAMREIKGIGVASVPTQNTVIERFSQKYPHISPIWVKAKNG---CVKWNKVP 99  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 265 SPISEDLAKIEKSGVEVVV-----SPLTCEAARSVCQHCYGSWLAHA 309  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 100 SEVGADRVANVAVKFKNGIIDMGATTVDLVVNGSYEGAILPGFMMVHSLFRG 159  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 310 SMVDLGEAVGIIA-AQSIGEPGQLNMTFTHT-----GGVTFGEVAQOVRKIDG 358  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 160 TAKLP-----LVEVKPADFVVGKDTTEENIRLGVNGS-VYAL 195  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 359 TVKLPRKLRKTRVTRHGDALYVEANGIMLEPTR--VGDVTPNQEVHLTGGSTLYVF 416  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 196 EGIIGRIKEVYGDLPVLTGGOS-----KIVKDM 224  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 417 DG--NKVNKQV-LLAEVALGGRTTRTNTTEKAVKDV 448  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC -----

RESULT 6  
 ID Y840.METJA STANDARD; PRT; 330 AA.  
 AC Q58250;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0840.  
 GN MJ0840.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uytterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: STRONG, TO M.VANNIELII HYPOTHETICAL PROTEIN IN ARGG  
 CC 5'REGION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U67528; AAB98845.1; -;  
 CC TIGR; MJ0840; -;  
 CC InterPro; IPR002756; DUF66.  
 CC Pfam; PF01897; DUF66; 1.  
 CC ProDom; PD024636; DUF66; 1.  
 CC Hypothetical protein; Complete proteome.  
 CC KW Hypothetical protein; Complete proteome.  
 CC SEQUENCE 330 AA; 37376 MW; 5A3EAA6D2125D829 CRC64;  
 CC -----  
 CC Query Match 7.2%; Score 92; DB 1; Length 330;  
 CC Best Local Similarity 20.4%; Pred. No. 1.5;  
 CC Matches 66; Conservative 51; Mismatches 106; Indels 100; Gaps 14;  
 CC -----  
 CC 1 MYLLVDVGNTHSVFSTE-DGKTRRWRLSTGVFQTEDELFSLHPLIGDAMREIKGIV 59  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 4 MILGIDIGGANT--KITEIGEDNYKIHIIYFPMKKKDEL-EDLLKNYNDVYVALVMT 60  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 60 ASVVPQNT-----VIERFSQKYPHISPIWVKAKNG---CVKWNKVPSEVGADRVAN 109  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 61 AELADCYKTKKEGVEDLIIDKV-EKAFN-CPYVFDVNGNFVTSEAKKNYLDVSASNNA 118  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 110 VVAFVREYKNG-IIDMGATT-----VDLVVNGS--YEGGAILPGFF 150  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 119 TAKFAVEFIKDCILVDMGSTTDDIIPKDKVELAKETDLDRLNNQLVYVGLTPTVSF 178  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 151 MWVHSLFRG-----TAKPLVEVKPADFVVGKDTTEENIRLGVNGSVYALEGI 198  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 179 LANKIEFRGLTNLSSEYFAITADISL-----ILNKITEEDYTCDDPDGAGKDFESC 230  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 199 IGRIKEV-----YGDLPVLTGGOSKI 220  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 231 LTRLVRLCADREWMKDELDIFANKLYNKLELIRENVDTIAKRYNLNDVVITGLGEI 290  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 221 VKDMIKH-----EIFDEDTI 236  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC 291 LKDALDEYNIISIKETYGKDVSL 313  
 CC ||||| : : : : : ||| : : : : : ||| : : : : :  
 CC -----

## RESULT 7

ID AROA.METJA STANDARD; PRT; 429 AA.  
 AC Q57925;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable 3-phosphohikimate 1-carboxyvinyltransferase (EC 2.5.1.19)  
 DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).  
 GN AROA OR MJ0502.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;



```

RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67500; AAB98493.1; -
CC TIGR; MJ0502; -
CC InterPro; IPR001986; EPSP_synthase.
CC Pfam; PF00275; EPSP_synthase; 1.
CC PRODOM; PD001867; EPSP_synthase; 1.
CC PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
CC KW Hypothetical protein; Aromatic amino acid biosynthesis; Transferase;
CC Complete proteome.
CC SEQUENCE 429 AA; 47008 MW; EELD8C083B934D59 CRC64;

Query Match 7.1%; Score 91.5; DB 1; Length 429;
Best Local Similarity 21.2%; Pred. No. 2.3;
Matches 53; Conservative 52; Mismatches 98; Indels 47; Gaps 13;

QY 4 LVVGNTHSVFSITED--GKTFRRWLSTGVQFTEDELFSH-LHPLLDGDMREIKGIGVA 60
DB 86 IIDIGNSGTTIRLITSTIASQIPKGYAILTG-----DDSRKRPMPQLL-DALKQLN----- 135
QY 61 SWPTQNTVIERFSOKYPHISPIWVKAKNGCVKN--VKNPSEVGADRVAN---VVAFVKE 116
DB 136 -----IEAFSKLDGTAPIIV--KSGKIYGNVVKIRGDISQFITSMLMLLPENKE 184
QY 117 YKNGIITDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPVLEVYKPADVVVG 176
DB 185 DTEIILTSPLKSPYIDITLD-----ILNFKGIKIDKTONGELVYGNQKYPIDYIVE 237
QY 177 KDTEN---IRLVVNGSVYALEGIIIGRIKEVYGDLPVV---LTVGGOSKIVKDMIKHEI 229
DB 238 GDYSASYSLIAAGVLINSNITENLFANSQ--GDKATINIVKMGADIKVKKDKV---I 292
QY 230 FDEDLTIKGV 239
DB 293 IEGEYSLKGI 302

RESULT 8
RPOC_LEUPE
ID RPOC_LEUPE STANDARD; PRT; 989 AA.
AC P94892;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit) (Fragment).
GN RPOC.

```

```

OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 523;
RX MEDLINE=97016803; PubMed=8863429;
RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;
RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does
RT not support the hypothesis inferred from 16S rRNA analysis that
RT oenococcus oeni (formerly leuconostoc oenos) is a tachytelic
RT (fast-evolving) bacterium."
RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X95810; CAA65077.1; -
CC HSP; Q9KWU6; 1HQM.
CC InterPro; IPR000722; RNA_pol_A.
CC InterPro; IPR002879; RNA_pol_A2.
CC Pfam; PF00623; RNA_pol_A; 1.
CC Pfam; PF01854; RNA_pol_A2; 1.
CC KW Transferase; DNA-directed RNA polymerase; Transcription.
CC NON_TER 1
CC NON_TER 989
CC SEQUENCE 989 AA; 110185 MW; 395E91BE46F43CDC CRC64;

Query Match 7.1%; Score 90.5; DB 1; Length 989;
Best Local Similarity 21.8%; Pred. No. 7.7;
Matches 57; Conservative 40; Mismatches 105; Indels 59; Gaps 12;

QY 6 DVGNTHSVFSTIEDGKTFRRWLSTGVQFTEDELFSHLLPLLDGDMREIKGIGVASVVP 65
DB 379 DLGYDKS-----TESGLTVS-----MTDVLELEK-----PALEDAHSQVA----- 615
QY 66 QNTVIERF-----SOKYFHISPIWVKAKN-----GCVKWVKNPSEVGADRVN---NVV 111
DB 516 --TVTKQFRGLITDSERYQVTEIWTAKADIITQDKLIESFEPTNPFMDQSGAGNIS 673
QY 112 AFVKEYGKNGIITDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLP----- 164
DB 574 NFVQLAGMGLMAGPG--GKLIPLVTANFREGITVNMETFTSHGARGMSDTALKTANS 732
QY 165 -----LVEYKPADVVY---GKDTENIRLVVNGSVYALEGIIIGRIKEVYGDLPVLTGG 216
DB 733 YLTFRLVDV--AQDVIVREFDNDSDRGVAAQIMDGTWSVEPLVYRILGRYAMKSVFDPET 791
QY 217 QSKIVKDMIKHEIFDEDLTIK 237
DB 792 DEKICS---RNEMIDEVAKK 809

RESULT 9
RPOC_LEUPS
ID RPOC_LEUPS STANDARD; PRT; 989 AA.
AC P94899;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
GN RPOC.

```

DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase  
 DE beta' chain) (RNA polymerase beta' subunit) (Fragment).  
 GN RPOC.  
 GN Leuconostoc pseudomesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=33968;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCDO 768;  
 RX MEDLINE=97016803; PubMed=8863429;  
 RA Morse R., Collins M.D., O'Hanlon K., Wallbanks S., Richardson P.T.;  
 RT "Analysis of the beta' subunit of DNA-dependent RNA polymerase does  
 RT not support the hypothesis inferred from 16S rRNA analysis that  
 RT Oenococcus oeni (formerly Leuconostoc oenos) is a tachytelic  
 RT (fast-evolving) bacterium."  
 RL Int. J. Syst. Bacteriol. 46:1004-1009(1996).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X95812; CAA65079.1;  
 DR HSP; Q9KW06; LHQM.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR Transferase; DNA-directed RNA polymerase; Transcription.  
 FT NON\_TER 1  
 FT SEQUENCE 989 989  
 SQ SEQUENCE 989 AA; 109914 MW; 55B1C8CA8739D06 CRC64;  
  
 Query Match 7.0%; Score 89.5; DB 1; Length 989;  
 Best Local Similarity 21.0%; Pred. No. 9.3;  
 Matches 58; Conservative 41; Mismatches 106; Indels 71; Gaps 12;  
  
 Db 6 DVGNTSHVTSITDGTFRWRRLSTGVFTQDELFSHLPLLDGAMREIKGIGVASVPT 65  
 579 DLGYEKS-----TESGLT-----VAMTDVTLKEKPAILEDHQA----- 615  
  
 Qy 66 QNTVIERP-----SOKTFHSPVWVAKN-----GCVKWNKPNSEVGADVA--NVV 111  
 616 --TVTKQFRGLTDDERYQRTVEITWKARDIIQDKLIESFEPTNPFMMQDSGARGNIS 673  
  
 Qy 112 AFVKEYGKGIIDMGTAITVDLVVNGSEGGAILPGFFMMVHSLPRTAKLP----- 164  
 674 NFVQLAGMGLMAGPG-GKIIPLVTVANFREGUTVMEMFISTHGARGMSDTALKANSG 732  
  
 Qy 165 -----LVEVKPADFVY---GKDFEENIRGLVNGSVVALEGIIGRIKEVYGDLPVVLTTG 216  
 733 YLTRLVDV-AQDVIVREFDNDSDRGVAKVAKMDGTSVVEPLDRLILGRVAMKSVDPET 791  
  
 Qy 217 QSKIVKDMIKHEIFDE-----LTIKGVY 240  
 792 GEKIVS---RNEMIDEDVAKAIVNAGIEETIRSVF 824

RESULT 10  
 NUAN\_ACACA

ID NUAN\_ACACA STANDARD; PRT; 675 AA.  
 AC Q37373;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE NADH dehydrogenase, subunit 11 (EC 1.6.5.3) (NADH dehydrogenase  
 DE (ubiquinone)) (ubiquinone reductase) (Type I dehydrogenase) (Complex I  
 DE dehydrogenase).  
 DE NAD11.  
 GN Acanthamoeba castellanii (Amoeba).  
 OS Mitochondrion.  
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 OX NCBI\_TaxID=5755;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 30010 / NEFF;  
 RX MEDLINE=95147275; PubMed=7844823;  
 RA Burger G., Planete I., Ionergan K.M., Gray M.W.;  
 RT "The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba  
 RT castellanii: complete sequence, gene content and genome  
 RT organization."  
 RL J. Mol. Biol. 245:522-537(1995).  
 CC -1- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A  
 CC COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT MAY  
 CC FORM PART OF THE ACTIVE SITE CREVICE WHERE NADH IS OXIDIZED (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- COFACTOR: BINDS 1 2FE-2S CLUSTER AND 1 4FE-4S CLUSTER (Potential).  
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE  
 CC MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 kDa SUBUNIT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; U12386; AAD11824.1;  
 DR InterPro; IPR000283; Complex1\_75K.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF00111; fer2; 1.  
 DR PROSITE; PS00641; COMPLEX1\_75K\_1; FALSE\_NEG.  
 DR PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
 DR PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
 DR Oxidoreductase; NAD; ubiquinone; Mitochondrion; Iron-sulfur; 4Fe-4S.  
 FT METAL 25 25 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 36 36 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 47 47 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 50 50 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 100 100 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 103 103 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 109 109 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 148 148 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 151 151 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 154 154 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 198 198 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 SQ SEQUENCE 675 AA; 77554 MW; 3870F3BC9E8B9807 CRC64;

Query Match 6.9%; Score 89; DB 1; Length 675;  
 Best Local Similarity 19.1%; Pred. No. 6.5;  
 Matches 54; Conservative 37; Mismatches 93; Indels 98; Gaps 11;  
  
 Qy 22 TFRWRRL-SRGVFTQDELFSHLPLLDGAMREIKGIGVASVPTQNTVIE-----R 72  
 Db 211 TSPWELKSYNSIDVLDLSHNSIRV-----DIRGTYKIMRLPRVNSLNDWTDKIR 263  
 Qy 73 FSQYTHISPIW---VKAKNGCVKWNKPNSEVGADRVANVAVFYKEY----- 117

```

Db 264 PSYDFRQRQLDPMVKISGSLKLGWK-----KAMLFKKFCNLFGRNHSF 312
QY 118 -----GKNGIIDIIMGTATTVD--LVYNGSYEGGAILPGFFMVSLSFRGTA---KLPLVE 167
Db 313 IPLRCYIGDLDLEIYIFKFLNGS-----NFFLPSSYNDLTALYSFNTPLTR 364
QY 168 VKPADVFVGKQTEENIRLGVNGSV-----192
Db 365 LDEGDFCILLDNLRLVLPFVNSRIKQLVSKKMLPVFLGFSYFNFYKHSNSKTL 424
QY 193 VALEG---IIGRIKEVGDLPVVLATGGQSKIIVKDMIKHEIF 231
Db 425 HVLEGSHWSAKISKFSKPIFLIGDSSSLKGLSLIVPLFN 466

RESULT 11
LE11_ARCFU
--LE11_ARCFU STANDARD; PRT; 489 AA.
O29305;
15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 2-isopropylmalate synthase 1 (EC 4.1.3.12) (Alpha-isopropylmalate
synthase 1) (Alpha-IPM synthetase 1).
GN LEU1 OR AF0957.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Catalyzes the condensation of the acetyl group of
acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
CC -1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA -
acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
CC -1- PATHWAY: Leucine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
SYNTHASE FAMILY. LEU1 A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF0957;
CC TIGR; AF0957;
CC InterPro; IPR002034; AIPM_Hcitt_synth.
CC InterPro; IPR000891; HMGL-like.
CC Pfam; PF00682; HMGL-like; 1.
CC PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
CC PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
CC Leucine biosynthesis; Lyase; Complete proteome.
KW

```

```

SQ SEQUENCE 489 AA; 53345 MW; 9E84B204DE796F84 CRC64;
Query Match . 6.9%; Score 88.5; DB 1; Length 489;
Best Local Similarity 22.5%; Pred. No. 4.8;
Matches 48; Conservative 37; Mismatches 91; Indels 37; Gaps 7;
QY 45 PLIGD-AMREIKGIVASVVPTQNTVIERFSQKVFHISPIWVKAKNGCVKWNVPNSVEG 103
Db 274 PIVGDNAFTHESGI-----HTSALFRDAKSVEPISPEVVGKRVYVLGKHAGRASVE 325
QY 104 A-----DRVANVAVFVKEYGKNG-IIDMGATTVDLVVNGSYEGGAILPGFFM 151
Db 326 AINNELGYKATPEQMKELARIKEIGDKGRVTDADVRTIETVLQIKRKKVLEDL-- 383
QY 152 MVHSLFRGTAFLVEVKPADVFVGKQTEENIRLGVNGSVYALEGIIGRIKEYVGDLPV 211
Db 384 ---AIFSGKNVPMASVK-----LKIDGQERIEAAGVGLPVDVDAINAIRRAIKEFADIKL 435
QY 212 V-----LTGQSKIVKDMIKHEIFDEDLTIKG 238
Db 436 VSYHVDATGCTDALVDVVVQLKDKNKIVTARG 468

RESULT 12
CH60_AQUAE
ID CH60_AQUAE STANDARD; PRT; 545 AA.
AC O67943;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR MOPA OR GROEL OR AQ_2200.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
Graham D.E., Overbeek R., Snead M.A., Keller M., Aujaay M., Huber R.,
Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RT Nature 392:353-358(1998).
RL Nature 392:353-358(1998).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000777; AAC07897.1;
CC HSSP; P06139; IGRU.
CC InterPro; IPR001844; Chaperin_Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC Pfam; PF00118; cpn60_TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
CC Chaperone; ATP-binding; Complete proteome.
CC SEQUENCE 545 AA; 58794 MW; 6CB190ED68BC8B9 CRC64;
Query Match . 6.9%; Score 88; DB 1; Length 545;

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.0638 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-9  
Perfect score: 1281  
Sequence: 1 MYLLVDVGNTHSVFSTEDG.....HEIFDELTJIKGVYHFCFGD 246

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1281	100.0	246	16 Q9WZY5	Q9WZY5 thermotoga
2	389.5	30.4	254	16 Q9KGH5	Q9KGH5 bacillus ha
3	378.5	29.5	259	16 Q8YAC5	Q8YAC5 listeria mo
4	377.5	29.5	259	16 Q92F54	Q92F54 listeria in
5	371.5	29.0	258	2 Q9F985	Q9F985 bacillus st
6	360.5	28.1	273	16 Q97EB4	Q97EB4 clostridium
7	359.5	28.1	261	16 Q9A6Z1	Q9A6Z1 caulobacter
8	342.5	26.7	259	16 Q8XHL5	Q8XHL5 clostridium
9	342.5	26.7	265	16 Q9X8N6	Q9X8N6 streptomyc
10	324.5	25.3	255	16 Q8R7M2	Q8R7M2 thermoaer
11	317.5	24.8	256	16 Q8RFE4	Q8RFE4 fusobacteri
12	289	22.6	274	16 Q9CD56	Q9CD56 mycobacteri
13	286	22.3	272	16 Q96282	Q96282 mycobacteri
14	285	22.2	273	16 Q83446	Q83446 treponema p
15	260	20.3	262	16 Q9RX54	Q9RX54 deinococcus
16	221	17.3	212	2 Q32514	Q32514 desulfovibr

17	204.5	16.0	224	16 Q98093	Q98093 mycoplasma
18	203	15.8	262	16 Q51477	Q51477 borrelia bu
19	193.5	15.1	229	16 Q67753	Q67753 aquifex ae
20	179	14.0	223	16 Q92KY6	Q92KY6 helicobacte
21	165	12.9	223	16 Q25533	Q25533 helicobacte
22	150.5	11.7	592	16 Q9JW17	Q9JW17 neisseria m
23	149.5	11.7	295	16 Q8Y2M4	Q8Y2M4 ralstonia s
24	146	11.4	592	16 Q9JXF1	Q9JXF1 neisseria m
25	145	11.3	248	16 Q9HWC1	Q9HWC1 pseudomonas
26	139.5	10.9	209	16 Q9PIA9	Q9PIA9 campylobact
27	138	10.8	257	16 P74045	P74045 synecocyst
28	124.5	9.7	276	16 Q8YQD7	Q8YQD7 anabaena sp
29	103	8.0	242	16 Q9PC14	Q9PC14 xylella fas
30	101	7.9	322	2 Q8VVB8	Q8VVB8 streptococc
31	100.5	7.8	313	16 Q97FW3	Q97FW3 clostridium
32	100	7.8	831	16 Q9KZP4	Q9KZP4 streptomyc
33	98	7.7	315	16 Q51771	Q51771 borrelia bu
34	98	7.7	327	2 Q9RE12	Q9RE12 lactobacill
35	98	7.7	431	16 Q8XNA5	Q8XNA5 clostridium
36	96.5	7.5	816	16 Q8RB00	Q8RB00 thermoanaer
37	95	7.4	713	10 Q9FHG5	Q9FHG5 arabidopsis
38	95	7.4	780	16 Q9KR17	Q9KR17 bacillus ha
39	94.5	7.4	287	16 Q9KR61	Q9KR61 vibrio chol
40	94	7.3	1335	2 Q9LA54	Q9LA54 escherichia
41	93	7.3	1335	2 Q9LA58	Q9LA58 escherichia
42	92.5	7.2	320	16 Q97RW8	Q97RW8 streptococc
43	92	7.2	222	2 Q93PC4	Q93PC4 microscilla
44	92	7.2	298	16 Q92LP9	Q92LP9 rhizobium m
45	92	7.2	552	16 Q98AX9	Q98AX9 rhizobium l

ALIGNMENTS

RESULT 1					
Q9WZY5					
ID	Q9WZY5	PRELIMINARY;	PRT;	246 AA.	
AC	Q9WZY5				
DT	01-NOV-1999 (TREMBlrel. 12, Created)				
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)				
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)				
DE	Hypothetical protein TW0883.				
GN	TM0883.				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109;				
RX	MEDLINE=9287316; PubMed=10360571;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,				
RA	Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,				
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RT	genome sequence of Thermotoga maritima."				
RL	Nature 399:323-329(1999).				
DR	EMBL; AE001754; AAD35964.1; -				
DR	TIGR; TW0883;				
DR	InterPro; IPR004619; Baf.				
DR	Pfam; PF03309; Bvg_acc_factor; 1.				
DR	TIGRfams; TIGR00671; baf; 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;				
	Query Match 100.0%; Score 1281; DB 16; Length 246;				
	Best Local Similarity 100.0%; Pred. No. 2.2e-102;				
	Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MYLLVDVGNTHSVFSTEDGKTRWRWLSGVFQTEDELFSHLHLLGDAMREIKGIVA 60				

```

Db 1 MYLLVDVGNTHSVFSTEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAMREIKGICVA 60
QY 61 SVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVANVAFVKEYGKN 120
Db 61 SVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVANVAFVKEYGKN 120
QY 121 GIIDMGCTATTVDLVVNGSEGGAILPGFFMMVHSLFRGTAKPLVVEVKPADVVGKDT 180
Db 121 GIIDMGCTATTVDLVVNGSEGGAILPGFFMMVHSLFRGTAKPLVVEVKPADVVGKDT 180
QY 181 ENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIVKMDIKHEIFDEDLTIKGVY 240
Db 181 ENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIVKMDIKHEIFDEDLTIKGVY 240
QY 241 HFCEGD 246
Db 241 HFCEGD 246

RESULT 2
9KGH5 PRELIMINARY; PRT; 254 AA.
AC Q9KGH5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein BH0086.
GN BH0086.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001507; BAB03805.1; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 30.4%; Score 389.5; DB 16; Length 254;
Best Local Similarity 35.6%; Pred. No. 1.4e-25;
Matches 90; Conservative 54; Mismatches 88; Indels 21; Gaps 8;

QY 1 MYLLVDVGNTHSVFSTEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAMRE 53
Db 1 MYLLVDVGNTHSVFSTEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAMRE 53
QY 54 IKGIGVASVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVAN 109
Db 54 IKGIGVASVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVAN 109
QY 110 VVAFVKEYGKNGIIIDMGCTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168
Db 110 VVAFVKEYGKNGIIIDMGCTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168
QY 114 AVAAIELYGYPAIVDFGTATTVCLINEKKQYAGGVITAPGIMISTEALYHRASKLPRIE 173
Db 114 AVAAIELYGYPAIVDFGTATTVCLINEKKQYAGGVITAPGIMISTEALYHRASKLPRIE 173
QY 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIV-KDMIK 226
Db 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIV-KDMIK 226
QY 174 AKPKQVGTNTIDSMOSGIFGYVQSDGVVKKMKQAQSEPRKVIATGGLAKLIGTSET 233
Db 174 AKPKQVGTNTIDSMOSGIFGYVQSDGVVKKMKQAQSEPRKVIATGGLAKLIGTSET 233
QY 227 HEIFDEDLTIKGV 239
Db 227 HEIFDEDLTIKGV 239
QY 234 IDVIDSFLTLKGL 246
Db 234 IDVIDSFLTLKGL 246

RESULT 3
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL591974; CAD00748.1; -
DR MEROPS; M41.009; -
DR ListiList; LMO00221; -
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRFAMS: TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 29.5%; Score 378.5; DB 16; Length 259;
Best Local Similarity 35.6%; Pred. No. 1.3e-24;
Matches 90; Conservative 52; Mismatches 90; Indels 21; Gaps 7;

QY 1 MYLLVDVGNTHSVFSTEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAM 51
Db 1 MYLLVDVGNTHSVFSTEDGKTRRRLSTGVFQTEDELFSLHPLLLGDAM 51
QY 52 REIKGIGVASVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVAN 109
Db 52 REIKGIGVASVYPTQNTVIERESQKFFHISPIWVKAKNGCVKNNVKNPSEVGADRVAN 109
QY 55 SDIQGIILISSVVPPIHMETMTCVRYFNIRPLIVGPGIKTG-LNLKVDNPREIGSDRIVN 113
Db 55 SDIQGIILISSVVPPIHMETMTCVRYFNIRPLIVGPGIKTG-LNLKVDNPREIGSDRIVN 113
QY 110 VVAFVKEYGKNGIIIDMGCTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168
Db 110 VVAFVKEYGKNGIIIDMGCTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLVEV 168
QY 114 AVAAIEYGYPAIVDFGTATTTCFYIDSGYVGATAPGIMISTEALYHRAAKLPVVDI 173
Db 114 AVAAIEYGYPAIVDFGTATTTCFYIDSGYVGATAPGIMISTEALYHRAAKLPVVDI 173
QY 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIVKMDIKH 227
Db 169 KPADFVVGKDTENIRLGVVNGSVYALEGIIIGRIKEVYGDPLVPLVTGGQSKIVKMDIKH 227
QY 174 AESQIIGKSTVSSMOAGIFGYGQCEGIIAEMKKNASPVVAVATGGLARMITEKSSA 233
Db 174 AESQIIGKSTVSSMOAGIFGYGQCEGIIAEMKKNASPVVAVATGGLARMITEKSSA 233
QY 228 -EIFDEDLTIKGV 239
Db 228 -EIFDEDLTIKGV 239
QY 234 VDILDPFLTLKGL 246
Db 234 VDILDPFLTLKGL 246

RESULT 4
Q92F54 PRELIMINARY; PRT; 259 AA.
AC Q92F54;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein lin0253.

```





```

Db 68 VEGVLISSVVENIMYSLEHMIRKFKINPLVVGPGIKTGI---NIKYDNPKEVGADRVN 124
QY 110 VVATVKEGKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVE- 167
Db 125 AVAAHEIYKRSIIIDFGTATTCFCAVRENGDYLGAICPGIKVSSEALFEKAAKLPRVEL 184
QY 168 VKPADFVVGRTENIRLGVVNGSVYALEGIIGRIKEYV---GDLP--VVLTGGSQKIV 222
Db 185 IKPA-YAICRNTSISSGIVYIGVQRYIVIRMEKELOEPEGEKEPLVATGGLAKLIS 243
QY 223 DMIKH-EIFDEDLTIKGV 239
Db 244 EEAKNVDVINFLEGL 261

```

## RESULT 7

```

Q9A621 ID Q9A621 PRELIMINARY; PRT; 261 AA.
AC Q9A621 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN Cc1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

```

```

- Query Match 28.1%; Score 359.5; DB 16; Length 261;
Best Local Similarity 35.5%; Pred. No. 5.7e-23;
Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;

```

```

QY 1 MYLLVDVGNTHSVFSTEDGKTRFRWLSTGVFTQTEDELFSHLHPLL---GDAMREIKGI 57
Db 2 MLLAIEGNTNTMFAIHGASVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRADAV 61
QY 58 GVASVPTQNTVIERFSQKYPHISPIWV--KAKNGCVKNVKNPSEVGADRVANVAFVK 115
Db 62 IISVVPSQSTFNLRNLSRRYFNPLVIGENAKLG-IDVRIKPSSEAGADRLVNAIGAM 120
QY 116 EYKNGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEV-KPA-D 172
Db 121 VYGPPLVVIDSGTATTFDVAADGAFEGGIIAGINLSQALHEAAKLPRIAIQRPAGN 180
QY 173 FVVGKDEENIRLGVVNGSVYALEGIIGRIKEYVGD-LPVVLTGGSQKIV---DMIKHE 228
Db 181 RIVGTDVTSAMQSGFWGYSISLIEGLVARIKARGEPTVIATGGVASFEGATDSIDH- 239
QY 229 IFDEDLTIKGV 239
Db 240 -FSDSLTIRGL 249

```

## RESULT 8

```

Q8XHL5 ID Q8XHL5 PRELIMINARY; PRT; 259 AA.
AC Q8XHL5 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003194; BAB02174.1; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

```

```

Query Match 26.7%; Score 342.5; DB 16; Length 259;
Best Local Similarity 37.6%; Pred. No. 1.6e-21;
Matches 96; Conservative 44; Mismatches 94; Indels 21; Gaps 9;

```

```

QY 1 MYLLVDVGNTHSVFSTEDGKTRFRWLSTGVFTQTEDELFSHLHPLLGDAM---REIKGI 57
Db 1 MILLIDVGNTHSVFSTEDGKTRFRWLSTGVFTQTEDELFSHLHPLLGDAM---REIKGI 60
QY 58 GVASVPTQNTVIERFSQKYPHISPIWV--KAKNGCVKNVKN--NPSEVGADRVANVAF 113
Db 61 IISVVPMIMHSLNEMVRKCFCKEPIVVGPGIKTGI---NIKYDNPKEVGADRVANVAA 117
QY 114 VKEYGKGIIDMGTTATVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKPLPVEVKPAD 172
Db 118 FEKHKPMIIDFGTATTCFCAITKTEKGYLGNICPGIQISADALFERAAKLPRILEKPK 177
QY 173 FVVGKDEENIRLGVVNGSVYALEGIIGRIKEYVGDL---PVVL-TGGQSKIV---KDM 224
Db 178 SVICKNTVTSQAGIIYIGKVEYIVKRMKEMMDLGEKEFPFLATGLAKLVYSETDV 237
QY 225 IKHEIFDEDLTIKGV 239
Db 238 I--DEVDRKLTLEGL 250

```

## RESULT 9

```

Q9X8N6 ID Q9X8N6 PRELIMINARY; PRT; 265 AA.
AC Q9X8N6 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein SCO3380.
GN SCO3380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RX MEDLINE-97000351; PubMed-8843436;  
RA Kinashi H., Hopwood D.A.;  
RA "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RL coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL049628; CAB40880.1; .  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; Baf; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;  
  
Query Match 26.7%; Score 342.5; DB 16; Length 265;  
Best Local Similarity 35.7%; Pred. No. 1.7e-21;  
Matches 92; Conservative 45; Mismatches 98; Indels 23; Gaps 9;  
  
QY 1 MYLLVDVGNTHSVFSGTDEGKTRFRRLSTGVQFQTEDEL-----FSLHPLPLGDAMRE- 53  
DB 1 MLTTDVGNTHTVLGLDGEDIVHWRISTDSSRTADELAVLLQGLMGHPLGLDGLDGD 60  
  
QY 54 IKGIGVASVPTQNTVIERFSOKYFHSPIWVKANGCVKWNV----KNPSEVGADRVAN 109  
DB 61 IDGAIACATVPSPVLHREVTTRYGDVPA-VLVEPG-VKTGVPILTDPKVEGADRIIN 118  
  
QY 110 VVAFVKEYCKNGIIDMGATVTDLV-VNGSEGGAILPGFFMNVHSLFRGTAKPLPLVEV 168  
DB 119 AVAAVELYGGPAIVVDVDFGATTFDVAVSARGEYIGGVIAPEISVVALGVKGAQLKIEV 178  
  
QY 169 KPADFVVGKDTENIRLVNGSVYALEGIIIRIKRYEGDLP-----VLTGGQSKIVKDM 224  
DB 179 APRPSVIGNTVEMQSGIYGFAGQVDGVNRMARELADDDPDVTVIATGGLAPVW--L 236  
  
QY 225 IKHEIFDEP-----LTITGV 239  
DB 237 GESSVIDEHEPWLTIMGL 254  
  
RESULT 10  
Q8R7M2 PRELIMINARY; PRT; 255 AA.  
AC Q8R7M2  
DT 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DE Putative transcriptional regulator, homologs of Bvg accessory factor.

GN TTE2381.  
OS Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
OX NCBI\_TaxID=119072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MB4T / JCM11007;  
RX MEDLINE-21992816; PubMed-11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome.";  
RL Genome Res. 12:689-700(2002).  
DR EMBL; AE013180; AM25520.1; .  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;  
  
Query Match 25.3%; Score 324.5; DB 16; Length 255;  
Best Local Similarity 35.7%; Pred. No. 5.7e-20;  
Matches 89; Conservative 46; Mismatches 101; Indels 13; Gaps 7;  
  
QY 1 MYLLVDVGNTHSVFSGTDEGKTRFRRLSTGVQFQTEDE---LFSHLPLPLGDAMREIKGI 57  
DB 1 MLTAFDVGNTNIVMGVFKGKLLHSFRISTDKNTYDEYGLVNLQIGYNGISLTDV 60  
  
QY 58 GVASVPTQNTVIERFSOKYFHSPIWV--KAKNGCVKWNK--NPSEVGADRVANVAF 113  
DB 61 IISVVPPMLNTLVNLSKIFPKTPIVGVGPKTGI--NIKVDNPKVEGADRVANVAA 117  
  
QY 114 VKEYCKNGIIDMGATVTDLV-VNGSEGGAILPGFFMNVHSLFRGTAKPLPLVEV 172  
DB 118 YELYGVPVIVDFGATTFCAISERGEYLGIIAPGLMSADALFORTAKPLKIDLT 177  
  
QY 173 FVVGKDTENIRLVNGSVYALEGIIIRIKRYEGDLPVLTGGQSKIVKDKIKH-EIF 230  
DB 178 TVINRTVASMOSGIIYGVHGMVDYIVTRMKGEFAPSAYVAVTGTFANMAESKIDTV 237  
  
QY 231 DEDLTITGV 239  
DB 238 NEMLTLEGL 246  
  
RESULT 11  
Q8RFE4 PRELIMINARY; PRT; 256 AA.  
AC Q8RFE4  
DT 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE Bvg accessory factor.  
GN FN0761.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 25586;  
RX MEDLINE-21886394; PubMed-11889109;  
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fongstein M., Kyrpides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RL nucleatum strain ATCC 25586.";  
DR EMBL; AE010386; AAL94957.1; .  
KW Complete proteome.  
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;  
  
Query Match 24.8%; Score 317.5; DB 16; Length 256;  
Best Local Similarity 32.9%; Pred. No. 2.3e-19;





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: June 24, 2003, 21:27:56 ; Search time 7.65383 Seconds  
(without alignments)  
1007.182 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPPLAVDIGNTTVLGL.....DETLRLGLVLSRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	7.9	505	1	US-09-041-075A-8
2	99	7.5	984	4	US-09-287-354-2
3	99	7.5	1189	4	US-09-287-354-4
4	92	6.9	1189	4	US-09-287-354-3
5	89.5	6.8	2152	4	US-09-036-987A-3
6	89.5	6.8	2152	4	US-09-370-700-3
7	89	6.7	3816	4	US-09-428-517-3
8	89	6.7	4150	4	US-09-428-517-2
9	87.5	6.6	145	1	US-07-956-700B-111
10	87.5	6.6	145	1	US-08-476-537-111
11	87.5	6.6	145	1	US-08-485-607-111
12	87.5	6.6	145	2	US-08-475-879-111
13	87.5	6.6	145	4	US-09-433-043B-111
14	87.5	6.6	182	1	US-08-611-107-2
15	87.5	6.6	182	2	US-08-422-560A-2
16	87.5	6.6	182	4	US-08-468-793-2
17	87.5	6.6	593	4	US-09-433-043B-122
18	86.5	6.5	872	2	US-08-844-057-2
19	86.5	6.5	872	4	US-09-006-730-2
20	85	6.4	579	4	US-09-325-932A-185
21	84.5	6.4	1026	1	US-08-194-290-7
22	84	6.3	4551	3	US-09-320-878-1
23	84	6.3	4613	4	US-09-105-537-31
24	84	6.3	11877	4	US-09-105-537-6
25	82.5	6.2	355	4	US-08-818-112-79
26	82.5	6.2	355	4	US-08-818-111-80
27	82.5	6.2	355	4	US-09-056-556-79

28	82.5	6.2	355	4	US-09-072-596-80
29	82.5	6.2	1026	2	US-08-614-377A-7
30	82.5	6.2	1026	4	US-09-142-648B-7
31	82	6.2	731	2	US-08-911-364-1
32	82	6.2	733	4	US-08-464-700-2
33	81	6.1	332	4	US-09-036-987A-15
34	81	6.1	332	4	US-09-370-700-15
35	80.5	6.1	5215	4	US-09-105-537-2
36	79.5	6.0	388	4	US-09-191-608-22
37	79.5	6.0	460	2	US-08-912-129A-50
38	79.5	6.0	490	2	US-08-912-129A-54
39	79.5	6.0	618	2	US-08-817-900-2
40	79.5	6.0	796	2	US-08-817-900-2
41	79.5	6.0	796	4	US-09-236-645-2
42	79	6.0	384	1	US-08-707-793A-5
43	79	6.0	384	1	US-08-707-792A-5
44	79	6.0	3491	2	US-07-642-734C-2
45	79	6.0	3491	3	US-08-439-009A-2

ALIGNMENTS

RESULT 1  
US-09-041-075A-8  
; Sequence 8, Application US/09041075A  
; Patent No. H002022  
; GENERAL INFORMATION:  
; APPLICANT: Heidler, Steven A  
; APPLICANT: Radding, Jeffrey A  
; TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI  
; FILE REFERENCE: X-11242 Sequence Lst  
; Patent No. H002022  
; CURRENT APPLICATION NUMBER: US/09/041,075A  
; CURRENT FILING DATE: 1998-03-10  
; PRIOR APPLICATION NUMBER: 60/043,591  
; PRIOR FILING DATE: 1997-04-15  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 505  
; TYPE: PRT  
; ORGANISM: Candida parapsilosis  
US-09-041-075A-8

Query Match	7.9%	Score 104;	DB 1;	Length 505;
Best Local Similarity	25.3%	Pred. No. 0.0045;		
Matches	46;	Conservative	21;	Mismatches 47;
				Indels 68;
				Gaps 9;
QY	1	MPAPPLAVDIGNTTVLGLADASGALHTHTWRTNREMLP	-----DDLALQL----	48
DB	106	LPALPIL-----TWLGFYFTSSKIPHSWKPASVKVLPAMETILYGDLSNVLATIT	157	
QY	49	-----HGLFTLAGAPIRAAVLSVAPPVGENYALAKRHFMDAFVSAENLP	97	
DB	158	NRGLDILGNLPGLEHFGAPFVVAIMEFFAPPTS-----LRSF-----AFAGYNNLF	206	
QY	98	DVTVEL-----DTPGSGADRLCNLFGEKYLGGLDYAVVVD	134	
DB	207	GVIQMFPAAPPYKNLYGLQAPVNTMHSPPGLG--RIDELLGYDVTMTGFSNSPVI-	263	
QY	135	FG 136		
DB	264	FG 265		

RESULT 2  
US-09-287-354-2  
; Sequence 2, Application US/09287354  
; Patent No. 6348348  
; GENERAL INFORMATION:  
; APPLICANT: THOMPSON, Catherine C.  
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN

FILE REFERENCE: Thompson-20263/0243435  
CURRENT APPLICATION NUMBER: US/09/287,354  
CURRENT FILING DATE: 1999-04-07  
EARLIER APPLICATION NUMBER: US 60/080,888  
EARLIER FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 984  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-287-354-2

Query Match 7.5%; Score 99; DB 4; Length 984;  
Best Local Similarity 25.6%; Pred. No. 0.046;  
Matches 69; Conservative 24; Mismatches 107; Indels 70; Gaps 13;

QY 2 PAFPLLAVIDGNTTTLGLA-----DASGALTH-----TWIRTNREMLPDDLA 45  
698 PLGPPQSSLSGTTFWEGFSWPELRPKSDGVSLLHRLGDEDTSRVENLAASLPLPEY 757  
46 LQLHGLFTLAGAPIPRAAVLSSVAPPVGENYALALKR-HFMIDAFVSAENLPDVTVELD 104  
758 CALHGKLNLA-SYLPPLGLALRPLEPQWAAAYGVSPhRGLGTLKNCVADVADLVSLVHAD 816  
QY 105 TPGSVGADRLCNLFGEAK-YLGLDYAVVDFGTSTNFDVVGRRRLGGLATGAQVSA 163  
DB 817 TP-----LPWHRQAQKDFLSGLD-----GE-----GLWSPGSQVST 847  
QY 164 DALFARAAKLPRI-----TLQAPETAIGNTVHALQSLGVLFGYAEVMDGLLRIRAE----- 215  
DB 848 VWHVFRQAQDAQRIIRFLQWCPAGA-----GALEPGAP-GSCYLDAGLRLRREMGVS 900  
QY 216 -----LPGEAVAVATGFSRTVQICQEI 239  
DB 901 CWTLLQAPGEAVLVPAGA-PHQVQGLVSTV 929

## RESULT 3

US-09-287-354-4

Sequence 4, Application US/09287354

Patent No. 6348348

GENERAL INFORMATION:

APPLICANT: THOMPSON, Catherine C.

TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN

FILE REFERENCE: Thompson-20263/0243435

CURRENT APPLICATION NUMBER: US/09/287,354

CURRENT FILING DATE: 1999-04-07

EARLIER APPLICATION NUMBER: US 60/080,888

EARLIER FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4

LENGTH: 1189

TYPE: PRT

ORGANISM: Homo sapiens

US-09-287-354-4

Query Match 7.5%; Score 99; DB 4; Length 1189;  
Best Local Similarity 25.6%; Pred. No. 0.062;  
Matches 69; Conservative 24; Mismatches 107; Indels 70; Gaps 13;

QY 2 PAFPLLAVIDGNTTTLGLA-----DASGALTH-----TWIRTNREMLPDDLA 45  
903 PLGPPQSSLSGTTFWEGFSWPELRPKSDGVSLLHRLGDEDTSRVENLAASLPLPEY 962  
46 LQLHGLFTLAGAPIPRAAVLSSVAPPVGENYALALKR-HFMIDAFVSAENLPDVTVELD 104  
963 CALHGKLNLA-SYLPPLGLALRPLEPQWAAAYGVSPhRGLGTLKNCVADVADLVSLVHAD 1021  
QY 105 TPGSVGADRLCNLFGEAK-YLGLDYAVVDFGTSTNFDVVGRRRLGGLATGAQVSA 163  
DB 1022 TP-----LPWHRQAQKDFLSGLD-----GE-----GLWSPGSQVST 1052

QY 164 DALFARAAKLPRI-----TLQAPETAIGNTVHALQSLGVLFGYAEVMDGLLRIRAE----- 215  
DB 1053 VWHVFRQAQDAQRIIRFLQWCPAGA-----GALEPGAP-GSCYLDAGLRLRREMGVS 1105  
QY 216 -----LPGEAVAVATGFSRTVQICQEI 239  
DB 1106 CWTLLQAPGEAVLVPAGA-PHQVQGLVSTV 1134

## RESULT 4

US-09-287-354-3

Sequence 3, Application US/09287354

Patent No. 6348348

GENERAL INFORMATION:

APPLICANT: THOMPSON, Catherine C.

TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN

FILE REFERENCE: Thompson-20263/0243435

CURRENT APPLICATION NUMBER: US/09/287,354

CURRENT FILING DATE: 1999-04-07

EARLIER APPLICATION NUMBER: US 60/080,888

EARLIER FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 1189

TYPE: PRT

ORGANISM: Homo sapiens

US-09-287-354-3

Query Match 6.9%; Score 92; DB 4; Length 1189;  
Best Local Similarity 25.2%; Pred. No. 0.38;  
Matches 68; Conservative 24; Mismatches 108; Indels 70; Gaps 13;

QY 2 PAFPLLAVIDGNTTTLGLA-----DASGALTH-----TWIRTNREMLPDDLA 45  
DB 903 PLGPPQSSLSGTTFWEGFSWPELRPKSDGVSLLHRLGDEDTSRVENLAASLPLPEY 962  
46 LQLHGLFTLAGAPIPRAAVLSSVAPPVGENYALALKR-HFMIDAFVSAENLPDVTVELD 104  
963 CALHGKLNLA-SYLPPLGLALRPLEPQWAAAYGVSPhRGLGTLKNCVADVADLVSLVHAD 1021  
QY 105 TPGSVGADRLCNLFGEAK-YLGLDYAVVDFGTSTNFDVVGRRRLGGLATGAQVSA 163  
DB 1022 TP-----LPWHRQAQKDFLSGLD-----GE-----GLWSPGSQVST 1052  
QY 164 DALFARAAKLPRI-----TLQAPETAIGNTVHALQSLGVLFGYAEVMDGLLRIRAE----- 215  
DB 1053 VWHVFRQAQDAQRIIRFLQWCPAGA-----GALEPGAP-GSCYLDAGLRLRREMGVS 1105  
QY 216 -----LPGEAVAVATGFSRTVQICQEI 239  
DB 1106 CWTLLQAPGEAVLVPAGA-PHQVQGLVSTV 1134

## RESULT 5

US-09-036-987A-3

Sequence 3, Application US/09036987A

Patent No. 6143526

GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.

APPLICANT: Broughton, Mary C.

APPLICANT: Crawford, Kathryn P.

APPLICANT: Madduri, Krishnamurthy

APPLICANT: Merlo, Donald J.

APPLICANT: Treadway, Patti J.

APPLICANT: Turner, Jan R.

APPLICANT: Waldron, Clive

TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dow Agrosciences LLC Patent Department



	Query Match	6.7%	Score 89;	DB 4;	Length 3816; .	
	Best Local Similarity	25.0%;	Pred. No. 5;			
	Matches	65;	Conservative 33;	Mismatches 102;	Indels 60;	Gaps 13
QY	14	TTTVLGLDASGALHTHWIRTNREMLP----	DDIALQLHGLFTLAGAGPIPAAAVLSSVA	69		
	:	:::	: :	:	:	:
	:	:	:	:	:	:
Dd	504	SSTGAVVDVWGSLVATRSVFEEHVAMVTGLDLSMAGSLAGR--	AAGGVVP--GVVSGYA	559		
	:	:	:	:	:	:
QY	70	PPVGENYALALKRH-----	FMDIAFAVAEENLPDVTVELDPGSYGADRLCNLFCA	120		
	:	:	:	:	:	:
Dd	560	PAEGRRVVFPFGQSOWGMGAAGLLDACPFVEAEACAAVLDP-	LTGWSLVLEVRGG	617		
	:	:	:	:	:	:
QY	121	EKYLGLD-----	YAVVVDFTSNTF-----	DVVGRRREL-----	GGI-LATGAQV	161

Db 618 EAVLGRVDVQPALWAVMVSLARTWRYGVBPAAVVGSHQGEIARACVAGLSLADGARV 677  
Qy 162 SADALFARAALPRITLOAPETAIGKNTVH-ALQSLGVFGYAEVMDGLLRIRAEPLGEA 220  
Db 678 ----VILRSRAIARI-----AGGGGMVSVSLPAGRVRTMLDTYGGRL-----S 716  
Qy 221 VAVATGGFSRTVQICQEID 240  
Db 717 VAVNGPSSTVVSQDAQALD 736

## RESULT 8

US-09-428-517-2  
; Sequence 2, Application US/09428517  
; Patent No. 6251636  
; GENERAL INFORMATION:  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Shah, Sanjay Krishnakant  
; APPLICANT: McDaniel, Robert  
; APPLICANT: Tang, Li  
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30062-20029.00  
; CURRENT APPLICATION NUMBER: US/09/428,517  
; CURRENT FILING DATE: 1999-10-28  
; EARLIER APPLICATION NUMBER: 60/120,254  
; EARLIER FILING DATE: 1999-02-16  
; EARLIER APPLICATION NUMBER: 60/106,100  
; EARLIER FILING DATE: 1998-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4150  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
; OTHER INFORMATION: Oleandolide PKS  
US-09-428-517-2

Query Match 6.7%; Score 89; DB 4; Length 4150;  
Best Local Similarity 23.9%; Pred. No. 5.7;  
Matches 62; Conservative 35; Mismatches 104; Indels 58; Gaps 12;  
Qy 14 TTVVLGLADASGALHTWRTNREMLP-----DDLALQLHLGFTLAGAPIRAAVLSVA 69  
Db 3013 SSTGAGVYDVGWSLVATRSVFHRAVMVGTDLDSMAGSLAGF--AAGGVVP--GWVSGVA 3068  
70 PPVGENYALAKRH-----PMIDAFVSAENLPDVTVELDTPGSVGNDRLCNLFGA 120  
3069 PAEGRRVVFPPGGSQWQMAAGLLDACPFARAEACAAVLDR--LTGWSLVEVLRG 3126  
Qy 121 EKVYLGGLD-----YAVVDFGTSTNF-----DVVGRGRFL-----GGI-LATGAQV 161  
Db 3127 EAVLGRVDVQPALWAVMVSLARTWRYGVBPAAVVGSHQGEIARACVAGLSLADGARV 3186  
Qy 162 SADALFARAALPRITLOAPETAIGKNTVHALQSLGVFGYAEVMDGLLRIRAEPLGEAV 221  
Db 3187 ----VILRSRAIARIAGGGGMVSVGLS-----AERVRTMLDTYGGRV---SV 3226  
Qy 222 AVATGGFSRTVQICQEID 240  
Db 3227 AAVNGPSSTVVSQDAQALD 3245

## RESULT 9

US-07-956-700B-111  
; Sequence 111, Application US/07956700B  
; Patent No. 5539092  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; TITLE OF INVENTION: Carboxylase

NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5539092th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/956,700B  
FILING DATE: 19921002  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Thomas E. No. 5539092thrup  
REGISTRATION NUMBER: 33,268  
REFERENCE/DOCKET NUMBER: ARCD:058  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-312-744-0090  
TELEFAX: 1-312-755-4489  
INFORMATION FOR SEQ ID NO: 111:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-07-956-700B-111  
Query Match 6.6%; Score 87.5; DB 1; Length 145;  
Best Local Similarity 29.4%; Pred. No. 0.047; 60; Indels 15; Gaps 3;  
Matches 37; Conservative 14; Mismatches 14; Indels 15; Gaps 3;  
Qy 120 AEKYLGLDYAVVYDFGTSTNFDVVGRRGRFLGGILATGAQVSDALFARAALPRITLQ 179  
Db 19 AEVTLKSDDFELTVRKAVGVNNSVVPVVTAPLSGVVSGLP-SAIPIVAHAAPSP----- 72  
Qy 180 APEVTAIGKNTVHALQSLGVFGYAEVMDGLLRIR-----RAELPGEAVAVATGGFSR 230  
Db 73 SPEPTGRAADHAVTSSGSGQCAKIIDQKLAEVASPMVGTFTYRAPAPGEAVEVEVGDRI 132  
Qy 231 TVQGIC 236  
Db 133 QGQIVC 138  
RESULT 10  
US-08-476-537-111  
; Sequence 111, Application US/08476537  
; Patent No. 5756290  
; GENERAL INFORMATION:  
; APPLICANT: Robert Haselkorn and Piotr Gornicki  
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
; TITLE OF INVENTION: Carboxylase  
; NUMBER OF SEQUENCES: 116  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: 321 No. 5756290th Clark Street  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII-DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,537

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/956,700
; FILING DATE: 10/21/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas E. No. 5756290thrup
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: ARCD:058
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-312-744-0090
; TELEFAX: 1-312-755-4489
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Peptide
; -08-476-537-111
;
; Query Match 6.6% Score 87.5; DB 1; Length 145;
; Best Local Similarity 29.4%; Pred. No. 0.047;
; Matches 37; Conservative 14; Mismatches 60; Indels 15; Gaps 3;
;
; QY 120 AKEYLGGDLYAVVVDFCTSTNFDVVGRRGREFLGGILATGAQVSDADALFARAALKLPRTTLQ 179
; DB 19 AEVTLKSDDFELTVRAKGVNNSVPPVTAPLSGVVGSLP-SAIPVHAHASP----- 72
; QY 180 APTATGKNTVHALQSLGFLVGYAEMVDGLLRI-----RAELPGEAVAVATGGFSR 230
; DB 73 SPEPGTSRAADHAVTSSGSGPGAKIIDQKLAEVASPMVGTFTYRAPAPGEAVFEVGDRI 132
; QY 231 TVOGIC 236
; DB 133 QGQTV 138
;
; RESULT 11
; US-08-485-607-111
; Sequence 111, Application US/08485607
; Patent No. 5792627
; GENERAL INFORMATION:
; APPLICANT: Robert Haselkorn and Piotr Gornicki
; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA
; TITLE OF INVENTION: Carboxylase
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5792627th Clark Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,607
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/956,700
; FILING DATE: 10/21/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas E. No. 5792627thrup
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: ARCD:058
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-312-744-0090
; TELEFAX: 1-312-755-4489

```



```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,560A
; FILING DATE: 14-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,700
; FILING DATE: 02-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:152/WIM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-418-3000
; TELEFAX: 512-474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-422-560A-2

Query Match 6.6%; Score 87.5; DB 2; Length 182;
Best Local Similarity 29.4%; Pred. No. 0.067;
Matches 37; Conservative 14; Mismatches 60; Indels 15; Gaps 3;

QY 120 AEKYLGLDYAVVVDFTSTNFDVVGRRFLGILATGAQVSADALEFAAKLPRTLQ 179
   || | : | : | : | : | : | : | : | : | : | : | : | : |
Db 21 AEVTLKSDDFELTVRKAVGVNNSVYPVVTAPLSGVGSLP-SAIPIVAHAAPSP----- 74

QY 180 APETAIGKNTVHALQSLVFGYAEVMDGLLRRI-----RAELPGEAVAVATGGFSR 230
   : || : || : || : || : || : || : || : || : || : || : ||
Db 75 SPEPTSRADHAVTSSGSGPQKLIIDOKLAEVASPMVGTFTYRAPAPGEAVFVEVGDRIR 134

QY 231 TVQGIC 236
   | : |
Db 135 QGQTV 140
```

Search completed: June 24, 2003, 21:49:21  
Job time : 9.65383 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.9453 Seconds  
(without alignments)  
2373.336 Million cell updates/sec

Title: US-09-813-453A-8  
Sequence: 1 MPAPFLAVDIGNTTVLGL.....DETLLRGLVLSRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications, AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep:\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep:\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep:\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep:\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1324	100.0	262	9	US-09-813-453A-8
2	437.5	33.0	254	9	US-09-813-453A-47
3	428	32.3	258	9	US-09-813-453A-2
4	415.5	31.4	258	9	US-09-813-453A-49
5	414.5	31.3	255	9	US-09-813-453A-7
6	409.5	30.9	265	9	US-09-813-453A-4
7	405.5	30.6	262	9	US-09-813-453A-45
8	404	30.5	272	9	US-09-712-363-276
9	396.5	29.9	260	9	US-09-813-453A-51
10	378.5	28.6	233	9	US-09-813-453A-17
11	377	28.5	256	9	US-09-813-453A-55
12	347.5	26.2	258	9	US-09-813-453A-6
13	337	25.5	250	9	US-09-813-453A-3
14	333.5	25.2	219	9	US-09-813-453A-57
15	271	20.5	212	9	US-09-813-453A-59
16	260	19.6	245	9	US-09-813-453A-9
17	253.5	19.1	257	9	US-09-813-453A-53
18	248	18.7	273	9	US-09-813-453A-10
19					Sequence 10, Appl

20	215	16.2	241	9	US-09-813-453A-63	Sequence 63, Appl
21	159	12.0	262	9	US-09-813-453A-11	Sequence 11, Appl
22	133.5	10.1	229	9	US-09-813-453A-12	Sequence 12, Appl
23	131	9.9	244	9	US-09-813-453A-41	Sequence 41, Appl
24	128.5	9.7	242	9	US-09-813-453A-65	Sequence 65, Appl
25	123.5	9.3	257	9	US-09-813-453A-13	Sequence 13, Appl
26	104	7.9	505	9	US-09-742-580-8	Sequence 8, Appl
27	104	7.9	505	9	US-09-742-581-8	Sequence 8, Appl
28	104	7.9	505	10	US-09-742-582-8	Sequence 8, Appl
29	103.5	7.8	267	9	US-09-813-453A-15	Sequence 15, Appl
30	101.5	7.7	460	9	US-09-813-453A-39	Sequence 39, Appl
31	100.5	7.6	592	9	US-09-813-453A-22	Sequence 22, Appl
32	99	7.5	984	9	US-10-024-368-2	Sequence 2, Appl
33	99	7.5	1189	9	US-10-024-368-4	Sequence 4, Appl
34	99	7.5	1189	9	US-10-122-013-17	Sequence 17, Appl
35	98	7.4	249	9	US-09-813-453A-70	Sequence 70, Appl
36	96.5	7.3	592	9	US-09-813-453A-43	Sequence 43, Appl
37	95	7.2	249	9	US-09-813-453A-61	Sequence 61, Appl
38	95	7.2	655	9	US-09-906-209-10	Sequence 10, Appl
39	93.5	7.1	317	12	US-10-043-238-1	Sequence 1, Appl
40	93.5	7.1	317	12	US-10-043-238-3	Sequence 3, Appl
41	92	6.9	1189	9	US-10-024-368-3	Sequence 3, Appl
42	91.5	6.9	346	9	US-10-213-878-12	Sequence 12, Appl
43	91.5	6.9	346	9	US-10-214-059-12	Sequence 12, Appl
44	91.5	6.9	872	10	US-09-815-242-13295	Sequence 13295, A
45	90.5	6.8	346	9	US-10-213-878-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1  
US-09-813-453A-8  
; Sequence 8, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Deinococcus radiopugnans  
US-09-813-453A-8

Query Match	100.0%	Score 1324;	DB 9;	Length 262;
Best Local Similarity	100.0%;	Pred. No. 4.7e-111;	Mismatches 0;	Indels 0; Gaps 0;
Matches 262;	Conservative 0;			
Qy	1	MPAPFLAVDIGNTTVLGLADASGALTWTWRTNREMLPDDLALQLHGLFTLAGAPIP	60	
Db	1	MPAPFLAVDIGNTTVLGLADASGALTWTWRTNREMLPDDLALQLHGLFTLAGAPIP	60	
Qy	61	RAAVLSSVAPPVGENYALAKRHFMDIAFAVSAENLPDVTVELDTFGSVGADRLCNLFGA	120	
Db	61	RAAVLSSVAPPVGENYALAKRHFMDIAFAVSAENLPDVTVELDTFGSVGADRLCNLFGA	120	
Qy	121	EKYLGLDYAVVVDFTSTNFDVGRGRFLGILATGAQVSADALFAARAKLPRITLQA	180	
Db	121	EKYLGLDYAVVVDFTSTNFDVGRGRFLGILATGAQVSADALFAARAKLPRITLQA	180	
Qy	181	PETAIGKTVHALQSLVFGYAEWVDGLLRRTAEPLGGEAVAVATGGSFTVVGICOEID	240	

[illegible]

RESULT 4  
US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match	31.4%;	Score 415.5;	DB 9;	Length 258;
Best Local Similarity	35.2%;	Pred. No. 1.3e-29;		
Matches	94;	Conservative 49;	Mismatches 99;	Indels 25; Gaps 6;
QY	6	LLAVDIGNTTTVLGLADASSALHTWRIKTRTNEMLPDDLALQLHGLFTLAGAPIP--RAA	63	
Db	2	IFVLVGNTNTVLGVYDGD-ELKHWRITFSKSTEDGYGMIKALLNHVGLQFSDIRGI	60	
QY	64	VLSVAPPYGENYALAKR-----HFMDIAFVSAENLPDVTVELDTPGSVGADRLCN---	116	
Db	61	LISSVVPPI-----MFALERMCLKYFIKPLIVPGGIKTGLDIKYDNPREVGADRIVNVA	116	
QY	117	---LFGAEKYLGLDYAVVVDGTSNTFNFDVVGRRFLGGLIATGANOVSADALFAAAKL	173	
Db	117	GHLHGPS-----LIIDFGTATYTCYINEHKYMGGAAPGIMISTEALFAAAKL	168	
QY	174	PRITLQAPETAIGKNTVTHALQSLGVFGYAEWMDGLLRRIRAEELPGSAVATGGFSRTVQ	233	



Db 169 PRIETARPDIIGKNTVSANQAGLIGYVQGVESIVSRMAKSKIPPVKVIATGGLAPLIA 228  
Qy 234 GICQEIYDYDETLTGLRGLVWASRSR 260  
Db 229 SESDIIDVDPFELTGLKLLYEKNT 255

RESULT 5  
US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 31.3%; Score 414.5; DB 9; Length 255;  
Best Local Similarity 36.8%; Pred. No. 1.6e-29;  
Matches 96; Conservative 50; Mismatches 102; Indels 13; Gaps 6;  
Qy 6 LLAVIDGNTTTLGLADASGALHTWRTNREMLPDDALQHLGLFTLAGAPIP--RAA 63  
Db 2 LLTIDVGNTHVTLGLFDGEDIVEH--WRISTOSRRTADELAVLQGLMGHPLGLDELGDC 60  
Qy 64 VLSSVAPP--VGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCN--LFG 119  
Db 61 IISVVPLTGLVLERLSLG---YFGRPLVPGIKTGMPIDQNDPREVGADRIINAVAG 117  
Qy 120 AEKYLGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAACLPRITLQ 179  
Db 118 YEKRTSL---LIIVDFGTATFDVNRKGEYCGAIAPGLVISTEALFQRAKSLPRVDII 174  
180 APETAIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGAEAVAVATGGFSRTVQICOEI 239  
175 RPSAIIARNTVNSMQGIYGVGLVDEIVTRMAESKDAPRVATGGLASLIAPESKTI 234  
Qy 240 DYDETLTGLRGLVWASRSR 260  
Db 235 EAVEEYLTLEGLRILYERNRE 255

RESULT 6  
US-09-813-453A-4  
; Sequence 4, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-09-813-453A-4

Query Match 30.9%; Score 409.5; DB 9; Length 265;  
Best Local Similarity 37.9%; Pred. No. 4.8e-29;  
Matches 97; Conservative 44; Mismatches 102; Indels 13; Gaps 4;  
Qy 6 LLAVIDGNTTTLGLADASGALHTWRTNREMLPDDALQHLGLFTL-----AGAP 58  
Db 2 LLTIDVGNTHVTLGLFDGEDIVEH--WRISTOSRRTADELAVLQGLMGHPLGLDELGDC 60  
Qy 59 IPRAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLF 118  
Db 61 IDGTAICATVPSVLHRELREVTTRYGYDPAVLVEPGVKTPILTDPHPREVGAADRIINAV 120  
Qy 119 GAERYLGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAACLPRITL 178  
Db 121 AAVELYGG--PAIVVDFGTATFDVARSARGEYIGGVIAPIGIEISVEALGVKAQLRKIEV 178  
Qy 179 QAPETAIGKNTVHALQSLVFGYAEWVDGLLRIRAEPLGAEAVAVATGGFSRTVQGI 235  
Db 179 APRSRVIGKNTVEAMQSGIVYGFAGQVGVVNNMARELADDDPDDVTVIATGGLAPWVGE 238  
Qy 236 QCEIDYDETLTGLRGL 251  
Db 239 SSVIDEHEPWLTLMLGL 254

RESULT 7  
US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 30.6%; Score 405.5; DB 9; Length 262;  
Best Local Similarity 31.7%; Pred. No. 1.1e-28;  
Matches 84; Conservative 60; Mismatches 104; Indels 17; Gaps 4;  
Qy 6 LLAVIDGNTTTLGLADASGALHTWRTNREMLPDDALQHLGLFTLAGAPIP--RAA 63  
Db 2 IFVLVDGNTNAVGLVFE--EGELRQHRMETDRHKTEDEYGMVLKQLLEHGLSFEDVKGI 60  
Qy 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCN-----L 117  
Db 61 IVSSVVPPIIMFALERMCERYFKIKPLVVGPIKTNKIFENPREVGADRIINAVAGIHL 120  
Qy 118 FGAERYLGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAACLPRIT 177  
Db 121 YGSP-----LIIVDFGTATTYCINEEKHYMGVITPGIMISAEALYSRAAKLPRIE 172

QY 178 LOAPEAIGKNTVHALQSLVFGVYAEVMDGLLRIRAEPLGCEAVAVATGFSRTVQICQ 237  
Db 173 ITRPSSVGNKNTVSAMQSLGIVGVQVEGIVKRMKEAKQEPKVIATGGLAKLISESN 232  
QY 238 EIDYDETTLRGLVLEWASRSEVR 262  
Db 233 VIDVDPFTLKLGLMYERNANLQ 257

## RESULT 8

US-09-712-363-276  
; Sequence 276, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712.363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 276  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-276

Query Match 30.5%; Score 404; DB 9; Length 272;  
Best Local Similarity 38.3%; Pred. No. 1.5e-28;  
Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;  
QY 6 LLAVIDGNTTTLGLADASGALTHT-----WRITNREMLPDDLALQLHGLFTLAGAPIP 60  
Db 2 LLAIDVRNTHTVVGL--LSGMKEHAKVVOQWRITSEVTADELALTDGLIGEDSERLT 59  
QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTFGSVG 110  
Db 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIPEGVRTGPIPLVDNPKVEG 109  
QY 111 ARLCNLFGEAKYLGGLDYAVVDFGTSTNFDVGRGRFLGGILATGAQVSADALFARA 170  
Db 110 ADRIVNCLAA--YDRFRKAAIVVDGSSICVDVVSAGGEFLGGAIAAGVQVSSDAAAARS 167  
QY 171 AKLPRTLOAPETAIGKNTVHALQSLVFGVYAEVMDGLLRIRAEPLG-----EAVAVAT 225  
Db 168 AALRRVELARPSVVGKNTVECMQAGAVFGAGLVGLVGRIREDSVGSFVDHDAIVAT 227  
QY 226 GGFSTRVQICQIEDYDETTLRGLVLEWASRSEVR 262  
Db 228 GHTAPLLPELHTVDHYDQHLTLQGLRLVFERNLEVQ 264

## RESULT 9

US-09-813-453A-5  
; Sequence 51, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813.453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-813-453A-5

Query Match 30.5%; Score 404; DB 9; Length 272;  
Best Local Similarity 38.3%; Pred. No. 1.5e-28;  
Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;  
QY 6 LLAVIDGNTTTLGLADASGALTHT-----WRITNREMLPDDLALQLHGLFTLAGAPIP 60  
Db 2 LLAIDVRNTHTVVGL--LSGMKEHAKVVOQWRITSEVTADELALTDGLIGEDSERLT 59  
QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTFGSVG 110  
Db 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIPEGVRTGPIPLVDNPKVEG 109  
QY 111 ARLCNLFGEAKYLGGLDYAVVDFGTSTNFDVGRGRFLGGILATGAQVSADALFARA 170  
Db 110 ADRIVNCLAA--YDRFRKAAIVVDGSSICVDVVSAGGEFLGGAIAAGVQVSSDAAAARS 167  
QY 171 AKLPRTLOAPETAIGKNTVHALQSLVFGVYAEVMDGLLRIRAEPLG-----EAVAVAT 225  
Db 168 AALRRVELARPSVVGKNTVECMQAGAVFGAGLVGLVGRIREDSVGSFVDHDAIVAT 227  
QY 226 GGFSTRVQICQIEDYDETTLRGLVLEWASRSEVR 262  
Db 228 GHTAPLLPELHTVDHYDQHLTLQGLRLVFERNLEVQ 264

## RESULT 10

US-09-813-453A-51  
; Sequence 51, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: CGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813.453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Caulobacter crescentus  
US-09-813-453A-51





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.55691 Seconds  
(without alignments)  
2384.688 Million cell updates/sec

Title: US-09-813-453a-8  
Perfect score: 1324  
Sequence: 1 MFAPPLAVIDIGNTTVLGL.....DETTLTURLVLMASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues 112892  
Jtal number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378.5	28.6	233	1 YACB_BACSU	P37564 bacillus su
2	116	8.8	336	1 Y757_STRCO	Q9j45 streptomyce
3	103.5	7.8	267	1 BAF_BOPE	Q45338 bordetella
4	102.5	7.7	675	1 ATKD_DEIRA	Q9zpo deinococcus
5	99	7.5	1189	1 HAIR_HUMAN	Q43593 homo sapien
6	95	7.2	872	1 SYA_STRPY	Q99257 streptococ
7	94.5	7.1	310	1 KHSF_THEAC	Q9hkr6 thermoplas
8	94.5	7.1	314	1 MDH_RICCN	Q92ia0 rickettsia
9	93.5	7.1	419	1 AROA_METH	Q26860 methanobact
10	91	6.9	439	1 AROA_HALN1	Q9hql halobacteri
11	89.5	6.8	292	1 DAPF_RALSO	Q9y344 ralstonia s
12	89	6.7	383	1 HIS2_NEIMA	Q9jv26 neisseria m
13	88.5	6.7	328	1 GPDA_DEIRA	Q9rr76 deinococcus
14	88.5	6.7	521	1 VL2_HPVO4	Q07862 human papil
15	88.5	6.7	1403	1 VG22_HSV11	Q00105 ictaluriid h
16	88	6.6	321	1 GLK_BACSU	P34495 bacillus su
17	87.5	6.6	181	1 BCCP_ANASP	Q06881 anabaena sp
18	87.5	6.6	704	1 DP3E_RHOCA	Q68045 rhodobacter
19	87.5	6.6	872	1 SYA_STRPN	Q97q48 streptococ
20	87	6.6	365	1 LPXK_CHLPN	Q28223 chlamydia p
21	87	6.6	383	1 HIS2_NEIMB	Q9K013 neisseria m
22	87	6.6	411	1 VGLM_HSVBC	P52370 bovine herp
23	86.5	6.5	526	1 VP5_BTVA2	P30209 bluetongue
24	85.5	6.5	218	1 Y5A4_STRCO	Q9tkm3 streptomyce
25	85	6.4	603	1 FTSH_CYAME	Q9tj83 cyanidiosch
26	84.5	6.4	928	1 DPOI_SALTY	Q9fi73 salmonella
27	84.5	6.4	1025	1 SLAP_CAUCR	P35828 caulobacter
28	84	6.3	360	1 LPXD_CHLPN	Q9z8n6 chlamydia p
29	84	6.3	482	1 PTSD_VIBCH	Q9kvd9 vibrio chol
30	83.5	6.3	348	1 LPXK_AGRF5	Q8uh15 agrobacteri
31	83.5	6.3	375	1 HEMZ_STRCO	Q50533 streptomyce
32	83.5	6.3	377	1 HIS8_MYCIE	Q9x7b8 mycobacteri
33	83	6.3	499	1 GLCD_ECOLI	P52075 escherichia

34	82.5	6.2	230	1 RNFE_SALTY	Q8xex9 salmonella
35	82.5	6.2	231	1 RNFE_ECOLI	P77179 escherichia
36	82.5	6.2	511	1 DLDH_SCHPO	O00087 schizosacch
37	82.5	6.2	1320	1 PUTA_ECOLI	P09546 escherichia
38	82	6.2	750	1 HEPH_HSV11	P10192 herpes simp
39	81.5	6.2	627	1 FTH1_SYNY3	Q55700 synecocyst
40	81.5	6.2	933	1 SLAP_CAMEE	P35827 campylobact
41	81	6.1	719	1 ATKD_HALN1	P57699 halobacteri
42	81	6.1	835	1 CC48_YEAST	P25694 saccharomyc
43	80.5	6.1	409	1 PEPT_SALTY	P26311 salmonella
44	80.5	6.1	581	1 PBP2_NEIGO	P08149 neisseria g
45	80.5	6.1	644	1 FTSH_ODOSI	P49825 odontella s

## ALIGNMENTS

## RESULT 1

YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RT Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.N., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter K., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutu R., Wedler E., Wedler H., Weltzenegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.





## RESULT 5

HAIR\_HUMAN  
ID HAIR\_HUMAN STANDARD; PRT; 1189 AA.  
AC O43593; OGNEL;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hairless protein.  
GN HR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM), AND VARIANT ALUNC ALA-1022.  
RX MEDLINE=98111413; PubMed=9445480;  
RA Ahmad W., ul Haque M.F., Brancolini V., Tsou H.C., Ul Haque S.,  
"A  
Lam H., Alta V.M., Owen J., Deblaquiere M., Frank J.,  
Cserhalmi-Friedman P.B., Leask A., McGrath J.A., Peacocke M.,  
Ahmad M., Ott J., Christiano A.M.;  
"Alopecia universalis associated with a mutation in the human hairless  
gene.";  
RT Science 279:720-724(1998).  
RL [2]  
RN [3]  
RP SEQUENCE FROM N.A. (LONG ISOFORM), REVISIONS TO 572 AND 774, AND  
RP TISSUE SPECIFICITY.  
RP TISSUP-Peripheral blood leukocytes, and Skin fibroblast;  
RX MEDLINE=99162400; PubMed=10051399;  
RA Ahmad W., Zlotogorski A., Panteleyev A.A., Lam H., Ahmad M.,  
ul Haque M.F., Abdallah H.M., Dragan L., Christiano A.M.;  
"Genomic organization of the human hairless gene (HR) and  
RT identification of a mutation underlying congenital atrichia in an Arab  
RT Palestinian family.";  
RL Genomics 56:141-148(1999).  
RN [4]  
RP SEQUENCE FROM N.A., VARIANT ALUNC ASP-1136, TISSUE SPECIFICITY, AND  
RP ALTERNATIVE SPLICING.  
RP TISSUP-Peripheral blood leukocytes, Brain, and Fetal brain;  
RX MEDLINE=98409496; PubMed=9736769;  
RA Cichon S., Anker M., Vogt I.R., Rohleder H., Putzstuck M., Hillmer A.,  
Farooq S.A., Al-Dhafri K.S., Ahmad M., Haque S., Rietschel M.,  
Propping P., Kruse R., Nothen M.M.;  
"Cloning, genomic organization, alternative transcripts and mutational  
RT analysis of the gene responsible for autosomal recessive universal  
RT congenital alopecia.";  
RL Hum. Mol. Genet. 7:1671-1679(1998).  
RN [4]  
RP VARIANT APL GLN-620.  
RX MEDLINE=98431781; PubMed=9758627;  
RA Ahmad W., Irvine A.D., Lam H., Buckley C., Bingham E.A.,  
Panteleyev A.A., Ahmad M., McGrath J.A., Christiano A.M.;  
"A missense mutation in the zinc-finger domain of the human hairless  
RT gene underlies congenital atrichia in a family of Irish travellers.";  
RL Am. J. Hum. Genet. 63:984-991(1998).  
CC -1- FUNCTION: MAY ACT AS A TRANSCRIPTION FACTOR THAT COULD ACT ON TO  
CC REGULATE ONE OF THE PHASES OF HAIR GROWTH.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms; a long form (shown  
CC here) and a short form; are produced by alternative splicing.  
CC -1- TISSUE SPECIFICITY: Strongest expression of both isoforms is seen  
CC in the small intestine, weaker expression in brain and colon, and  
CC trace expression is found in liver, pancreas, spleen, thymus,  
CC stomach, salivary gland, appendix and trachea. Long isoform is  
CC always the most abundant. Long isoform is exclusively expressed at  
CC low levels in kidney and testis and short isoform exclusively at  
CC high levels in the skin.  
CC -1- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ALOPECIA UNIVERSALIS  
CC (ALUNC); A RARE AUTOSOMAL RECESSIVE FORM OF HAIR LOSS. IT IS  
CC CHARACTERIZED BY HAIR FOLLICLES WITHOUT HAIR.  
CC -1- DISEASE: DEFECTS IN HR ARE THE CAUSE OF ATRICHIA WITH PAPULAR  
CC LESIONS (APL) (ALSO KNOWN AS CONGENITAL ATRICHIA). THIS AUTOSOMAL

RECESSIVE DISEASE IS CHARACTERIZED BY PAPILLARY LESIONS OVER MOST  
OF THE BODY AND ALMOST COMPLETE ABSENCE OF HAIR.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF039196; AAC32258.2;  
DR EMBL; AJ277249; CAB87577.2;  
DR EMBL; AJ277250; CAB87577.2; JOINED.  
DR EMBL; AJ277251; CAB87577.2; JOINED.  
DR EMBL; AJ277252; CAB87577.2; JOINED.  
DR EMBL; AJ277253; CAB87577.2; JOINED.  
DR EMBL; AJ400825; CAB87577.2; JOINED.  
DR EMBL; AJ400826; CAB87577.2; JOINED.  
DR EMBL; AJ400827; CAB87577.2; JOINED.  
DR EMBL; AJ400828; CAB87577.2; JOINED.  
DR EMBL; AJ400829; CAB87577.2; JOINED.  
DR EMBL; AJ400830; CAB87577.2; JOINED.  
DR EMBL; AJ400831; CAB87577.2; JOINED.  
DR EMBL; AJ400832; CAB87577.2; JOINED.  
DR EMBL; AJ400833; CAB87577.2; JOINED.  
DR EMBL; AJ400834; CAB87577.2; JOINED.  
DR EMBL; AJ400835; CAB87577.2; JOINED.  
DR EMBL; AJ400836; CAB87577.2; JOINED.  
DR EMBL; AJ400837; CAB87577.2; JOINED.  
DR EMBL; AJ277165; CAB86602.1;  
DR Genew; HGNC:5172; HR.  
DR MIM; 602302; -.  
DR MIM; 203655; -.  
DR MIM; 209500; -.  
DR InterPro; IPR003347; TF\_JmJc.  
DR Pfam; PF02373; JmJc; 1.  
KW Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;  
KW Metal-binding; Alternative splicing; Disease mutation.  
FT ZN\_FING 600 625  
FT VARSPLIC 1072 1126 MISSING (IN SHORT ISOFORM).  
FT VARIANT 620 620 R -> Q (IN APL).  
FT VARIANT 1022 1022 /FTID-VAR\_005265.  
FT VARIANT 1136 1136 T -> A (IN ALUNC).  
FT VARIANT 1136 1136 V -> D (IN ALUNC).  
FT CONFLICT 337 337 D -> G (IN REF. 3).  
FT CONFLICT 446 446 L -> W (IN REF. 3).  
FT CONFLICT 584 584 S -> G (IN REF. 3).  
SQ SEQUENCE 1189 AA; 127509 MW; 5E244858716EB5DF CRC64;  
Query Match 7.5%; Score 99; DB 1; Length 1189;  
Best Local Similarity 25.6%; Pred. No. 1.5; Mismatches 107; Gaps 13;  
Matches 69; Conservative 24;  
Oy 2 PAFLLAVDIGNTTTVLGLA-----DASGALTH-----TWIRFTNREMLPDDL 45  
Db 903 PLGPPQSSLSGTTTFWEGFSWPELRPKSDGSLVLLHRLALGDEDTSRVENLAASLPPEY 962  
Oy 46 LQLHGLFTLAGAIPRAVLSSVAPPVGVENALAKR-HFMTDAFAVSAENLPDVTVELD 104  
Db 963 CALHGKLNLA-SYLLPGLALRPLEPQLWAAAYGVSPHRGLGTFKLCVEADLVSLVHAD 1021  
Oy 105 TPGSVGADRLCNLFGAEK-YLGGILDYAVVVDFTSTNFDVVGRRFLGILATGAQVSA 163  
Db 1022 TP-----LPAWRAQKDFLSGLD-----GE-----GLMSPGSGVST 1052  
Oy 164 DALFAAAKLPRI-----TLQAPETAIGKNTVHALQSGLVFGVYAEVMDGLLRIRAE---- 215  
Db 1053 VWHYFRAQDAQRIIRFLQMVCPAGN-----GALEPGAP-GSCYLDAGLRLRLREMGVS 1105  
Oy 216 -----LPGEAVAVATGTFSTVGOICQEI 239



Db	1106	CWTLQAPGEAVLPVAG-PRQVGLVSTV	1134		1		
Db	177	TLQAPETAIGNKNTVHALQSLVGYAEVMDGLLRIRARIELPGEAVAVATGGRFTVQ	234		1		
Db	723	TVKAPQLKDAAKVAQVLSDSL	773		1		
Qy	235	-----ICQEIYIDETLIRGLVLEWASR	258		1		
Db	774	GVRFASQVDVADAG-ALRTFADNWKOK	800		1		
RESULT 7							
ID	SYA_STRPY	STANDARD:	PRT:	872	AA.		
DT	15-JUN-2002	(Rel. 41, Created)					
DT	15-JUN-2002	(Rel. 41, Last sequence update)					
DT	15-JUN-2002	(Rel. 41, Last annotation update)					
DE	Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).						
GN	ALAS OR SPY1389						
OS	Streptococcus pyogenes.						
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;						
OC	Streptococcus.						
OX	NCBI_TaxID=13114;						
RN	[1]						
SEQUENCE FROM N.A.							
STRAIN=SF370 / ATCC 700294 / Serotype M1;							
MEDLINE=21192684; PubMed=11296296;							
Ra	Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,						
Ra	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,						
Ra	Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,						
Ra	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.						
RT	"Complete genome sequence of an M1 strain of Streptococcus pyogenes."						
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).						
CC	- - CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) -> AMP +						
CC	diphosphate + L-alanyl-tRNA(Ala).						
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.						
CC	- - SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/						
CC	or send an email to license@isb-sib.ch).						
CC	-----						
EMBL; AB006576; AAK34208.1;							
DR	InterPro; IPR002106; AATRNA_ligaseII.						
DR	InterPro; IPR003156; DHHA1.						
DR	InterPro; IPR002318; tRNA-synt_2c.						
DR	Pfam; PF02272; DHHA1.1.						
DR	Pfam; PF01411; tRNA-synt_2c; 1.						
DR	PRINTS; PR00980; TRNASYNTHALA.						
DR	TIGRFAMs; TIGR00344; alas; 1.						
DR	PROSITE; PS50860; AA_TRNA_LIGASE_II_ALA; 1.						
SW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;						
KW	Complete proteome.						
SC	SEQUENCE 872 AA; 96530 MW; E4605AB8B195B95959 CRC64;						
Query Match	7.2%;	Score 95;	DB 1;	Length 872;			
Best Local Similarity	23.5%;	Pred. No. 2.3;					
Matches	77;	Conservative 35;	Mismatches 98;	Indels 118;	Gaps 17;		
Qy	21	ADASGALFTWTRIRTNREMLPDDALQLHGLTLAGAPIPRAVLSSVAPPVGVNVALAL 80			1		
Db	501	AEMGGVADTGRIKNDK---GDTVAEVVDVQKAPNGPLHTVNVLASLS--VGTNYTLEI 555			1		
Qy	81	KR-----HFMDIAF-----AVSAENLPDV 99			1		
Db	556	NKERRLAVEKNHTATHLLHAALHNVI GEHATQAGSLNEEFLEDFTHFEAVSNEELRHI 615			1		
Qy	100	TVE-----LDTPGSGVADRLCNLFGAEKY-----LGLGLYAVVVD 134			1		
Db	616	EQEVNQIWNALITTTTETDVTAKEMGA---MALFG-EKYGKVVVVQIG--NYSVELC 669			1		
Qy	135	FGTSTN-----FDVV-----GGRRFGLGILATGAOV-----SADALFARAAPKUPRI 176			1		
Db	670	GGTHLNSSEIGLFKIVKEEGIGSGTRRI--IAVTGRQAFYARNOEDALKEIAA----- 722			1		

```

QY 181 PETAIGNKNTVHALQSLVFGVAEMV-----DGLLR--RI-----RAELPGEAVAV 223
Db 196 YSKALGRTA--SLIAGLMSGNRLIRVGMNDIIVEPSRISLFPYYDMKRMALANEA VAA 253
QY 224 ATGFSRTVQICQEIYDETLRL-CLVELWA 256
Db 254 AVSAGPSILMVCDMSMD---SIRAGISEIFS 284

RESULT 8
MDH_RICCN
ID MDH_RICCN STANDARD; PRT; 314 AA.
AC Q92IAO;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH OR RC0520.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OC NCBI_TaxID=781;
OC [1]
RN STRAIN=Malish 7;
RP SEQUENCE FROM N.A.
RC MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxaloacetate + NADH.
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008614; AL03058.1; -
CC InterPro; IPR001252; Mdh.
CC Pfam; PF00056; ldh; 1.
CC Pfam; PF02866; ldh.C; 1.
CC PROSITE; PS00068; MDH; FALSE_NEG.
CC Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
FT ACT_SITE 150 PROTON-RELAY (BY SIMILARITY).
FT BINDING 153 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 177 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 314 AA; 33638 MW; 27F016CDE3EF6F14 CRC64;

Query Match 7.1%; Score 94.5; DB 1; Length 314;
Best Local Similarity 23.3%; Pred. No. 0.78;
Matches 74; Conservative 42; Mismatches 96; Indels 105; Gaps 15;

QY 16 TVLGLDASGALHTWRIR-----TNEMLPDDLALQHLGLFTLAG----- 56
Db 8 SLIGSGNIGTGTLAHLISRLSGDIVLFDVTGVPQGGKALDLMQAGTIAGSDIKIKGTNDY 67

QY 57 -----APIPRAAVLS-----SVAPPVGENYALALKRHFMDAFVSAENLPDV 99
Db 68 KDIEGSDAIIITAGLPKPKGMSREDLISINTGIMTKVAANKK-YAPDAFVIVITNPLDV 126

QY 100 TV-----ELDTP-----GSYGADRLCLNLGAKEY-----LGLDLYAVY--- 132
Db 127 MYVYMLKESGLPHKNYIGMAGVLDSSRF-NFLAEAEKSVNNVNSVNLGSHGDGMVPLA 185

QY 133 -----VDGTSTNFD---VVGRRPLGILA---TGAQVSADALFARAA--- 171

```

```

Db 186 RYSTISGVPIPDILIMGLSSNENIEKIIDTRNGGGEIVALKTGS-----AYYAPASA 240
QY 172 -----KLPRITIQAPETAIGKNTVHALQSLVFGVAEMVDGLLRRIRAEPLGSAVAV 223
Db 241 IEMLESYLKDKRQILITCAHLQGEVGHDLVGVPIIMIGK--EGLRVIELQLTAEKAL 298
QY 224 ATGFSRTVQICQEI 240
Db 299 ----FDRSVEGVKKLIE 311

RESULT 9
ARO_A_METH
ID ARO_A_METH STANDARD; PRT; 419 AA.
AC O26860;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
DE (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR MTH766.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC NCBI_TaxID=187420;
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Delta H;
RC MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumma W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate -
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000855; AAB85269.1; -
CC InterPro; IPR001986; EPSP_synthase.
CC Pfam; PF00275; EPSP_synthase; 1.
CC ProDom; PD001867; EPSP_synthase; 1.
CC PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE; PS00883; EPSP_SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 419 AA; 44747 MW; 395BDBB8C5262AC9 CRC64;

Query Match 7.1%; Score 93.5; DB 1; Length 419;
Best Local Similarity 25.5%; Pred. No. 1.3;
Matches 70; Conservative 38; Mismatches 101; Indels 65; Gaps 14;

QY 11 IGNT-----TTVLGLDASGALHTWRIRTNREMLP-----DDLALQHLGLFT 53
Db 87 LGNSGTTLRMTSVAGLAENYTVLTGDESLET-RPMQPLDLLALPLGVLEASSRMNGLPP 145

QY 54 L-----AGAPIPRAAVLS-----VAPPVGENYALALKRHFMDAFVSAENLPDV 100

```

```

Db 146 IIVRGGLRGSTIRSQDVSSQFISSILLIAPLTEGVVMEVGFISRPVY-----DMT 198
QY 101 VELDTPGSGADRCLNLFAGKYL-CGLDYAVVVDFGTSTNFDVVGRRFLGILATGA 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 VDMERFSVPDYSEGTFRVPAVYRGLDYTVBGDYSSAS-----YLAGAVAAAGG 249
QY 160 QVVSADALFAAKLPRITL-----QAPETAIGNTVHALQSLGVFGYAEVVDGLLRIRA 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 DVLIENLFRDSQGRDIIDISDMGAEVRRGSDHRIASTGELSG-----VSNLHDAPD 305
QY 215 ELPGAV--AVATGFGFSRT--VQIGCOEIDYDET 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 LLPTAVLALATG---RTEIGGV--EHARYKET 334

RESULT 10
ID AROA_HALN1 STANDARD; PRT; 439 AA.
AC Q9HQCL;
T 15-JUN-2002 (Rel. 41, Created)
F 15-JUN-2002 (Rel. 41, Last sequence update)
JT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable 3-phosphohikimate 1-carboxyvinyltransferase (EC 2.5.1.19)
DE (5-encolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR PSC OR VNGI2326.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lesky S.R., Balliga N.S., Thorsson V., Shroqna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitthausser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohikimate -
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphohikimate.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE005049; AAC19594.1;
CC InterPro: IPR001986; EPSP_synthase.
CC Pfam: PF00275; EPSP_synthase; 1.
CC ProDom: PD001867; EPSP_synthase; 2.
CC PROSITE: PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE: PS00885; EPSP_SYNTHASE_2; 1.
CC Aromatic amino acid biosynthesis; Transferase; Complete proteome.
CC SEQUENCE 439 AA; 44333 MW; 48CAD75A1F0EB89 CRC64;

Query Match 6.9%; Score 91; DB 1; Length 439;
Best Local Similarity 23.4%; Pred. No. 2.3;
Matches 64; Conservative 31; Mismatches 85; Indels 94; Gaps 13;

QY 2 PAFPLAVDIGNTTTVLGLADASGALTHTWRTNREMLPDDLALQLHG--LFTIAG--- 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 87 PAIPDAVIDCANGSTTMRVLVTAALADGTTVLTDGE-----SLRAPHGPLLDALSLGG 142
QY 57 -----APTPRAVLS--SVAPP--VGENYALAKRHEMIDAFVSAENLP----- 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 TARSTRNGQAPLVDPGVPVSGGSVALPGDVSSOFFVTAL-----LMAGAVTGTETDLTT 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 98 -----DVTVELDTPGSGADRCLC---NLFGAEKYL-CGLDYAVVVDFGTSTNFDVVG 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 ELKSAPYVDITLDVDFGASETAAGYRVRGQAYAPSGAEYAVPGDFSSAY----- 252
QY 146 RGRFLGILATGAQVSDALFAAKLPRITLQAPETAIGNTVHALQSLGVFGYAEV 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 -----LLAAGALAAADG-----AAVVVEGMHPSAQ-----GDRAIV 283
QY 206 D-----GLLRIRAEPLGEAVAVA 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 DVLRMGADIDWDTESGVITVQSELSGVEVGVA 317

RESULT 11
ID DAPF_RALSO STANDARD; PRT; 292 AA.
AC Q8Y344;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Diaminopimelate epimerase (EC 5.1.1.7) (DAP epimerase).
GN DAPF OR RSC0137 OR RS01049.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- CATALYTIC ACTIVITY: LL-2,6-diaminoheptanedioate - meso-
CC gammaaminoheptanedioate.
CC -!- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF LYSINE FROM ASPARTATE
CC SEMIALDEHYDE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DIAMINOPIMELATE EPIMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL646057; CAD13665.1;
CC InterPro: IPR001653; DAP_epimerase.
CC Pfam: PF01678; DAP_epimerase; 2.
CC TIGRFAMs: TIGR00652; DapF; 1.
CC PROSITE: PS01326; DAP_EPIMERASE;
CC Isomerase; Lysine biosynthesis; Complete proteome.
CC ACT_SITE 75 75 BY SIMILARITY.
CC ACT_SITE 226 226 BY SIMILARITY.
CC SEQUENCE 292 AA; 31628 MW; BB6CC03A2939CC51 CRC64;

Query Match 6.8%; Score 89.5; DB 1; Length 292;
Best Local Similarity 23.0%; Pred. No. 1.9;
Matches 65; Conservative 23; Mismatches 103; Indels 91; Gaps 13;

```

QY 47 QHGLFTLAGAPIPRAAVLSVAPPV---ENVALAKRHFMDA---FAVSAENLPDVT 100  
 Db 7 KMHG-----AGNDF---VLDGIATPIDFTPEQWRAIADRHFGVAGADQLLVERSTRDPVD 59  
 QY 101 VELDTGSGADRLCNLFNGAEKYLGLLDYAVVDFGTSTNFDVVGGRK-----RFLG 152  
 Db 60 FRY-----RIFNHDGGEVCECGNGARCFVKF-----VTDRLGDKTRIRVEVMN 103  
 QY 153 GI-----LATGAQVSAD---ALFAARAKLPRITLQAPETAIGNKNTVHALQ----- 194  
 Db 104 GISTLTMPQGVTVDMGAPVFAARLPFPVPDALPTRVEGRDQTHALQINGRTAWLSTVS 163  
 QY 195 -----SGLVF-----GYAEMVD---GLLRIRIAELPGEA 220  
 Db 164 MGNPHAVQVVDDAEAFVREDGPLIESHAVFPFRVNAAGFMEIADRAIRLVRVYERGAGET 223  
 QY 221 VAVATGGFSTVGGICQEIYDDETLTLRG--LVLEWASRSE 260  
 Db 224 LACGTGACAAVAGIRGLLDSPVKVTHGGDITIAWAGEGE 265

RESULT 12  
 HISZ\_NEIMA  
 ID HISZ\_NEIMA STANDARD; PRT; 383 AA.  
 AC Q9J726;  
 DT 13-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE ATP phosphoribosyltransferase regulatory subunit.  
 GN HISZ OR NMA1023.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;  
 RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.";  
 RT Nature 404:502-506(2000).  
 RL CC -!- FUNCTION: May allow the regulation of ATP phosphoribosyltransferase activity by histidine (By similarity).  
 CC -!- PATHWAY: Histidine biosynthesis; first step. Very important in the regulation of histidine metabolism.  
 CC -!- SUBUNIT: Binds to hisG (By similarity).  
 CC -!- MISCELLANEOUS: This function is generally fulfilled by the C-terminal part of hisG, which is missing in some bacteria such as this one.  
 CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC HISZ SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AL162754; CAB84292.1;  
 DR Histidine biosynthesis; Complete proteome.  
 KW SEQUENCE 383 AA; 41837 MW; B7BCA7266045F860 CRC64;  
 SQ  
 Query Match 6.7%; Score 89; DB 1; Length 383;  
 Best Local Similarity 22.9%; Pred. NO. 2.8;  
 Matches 54; Conservative 29; Mismatches 85; Indels 68; Gaps 12;

QY 9 VDIGNTTTVLGLADAGALHTWRTNREMLPDLALQLHGLFTLAGAPIP----- 60  
 Db 176 LDAGOSATLLALMOKDKTETVEAQVK-----AWKLDGMWAKAFSLLPRLYGGREV 225  
 QY 61 -----RAAVLSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTGSGADRL 114  
 Db 226 LSDARGRLPDLASVGGALGELQAVC-----DAF-----PDEIHD----- 261  
 QY 115 CNLFGAEKYLGLLDYAV-VVDFGTSTNFDVVGGRRE--LGGIL-----ATGAQVSADAL 166  
 Db 262 LSELVDNYHTGLLYAAYADF-----HDAVARGRVDGLGYYGRARPATGFSFDLSRF 316  
 QY 167 FARAANKL---PRITLQAPETAIGNKNTVHALQSLGVFGYAEVMD--GLLRIRIAELPFG 218  
 Db 317 IGRLPATERQPAVLVDAEDAEAAHEAVEALREQ---GQCVVDYIGIGHNVSEELAG 369

RESULT 13  
 GPDA\_DEIRA  
 ID GPDA\_DEIRA STANDARD; PRT; 328 AA.  
 AC Q9RR76;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 18-OCT-2001 (Rel. 40, Last sequence update)  
 DE Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-dependent glycerol-3-phosphate dehydrogenase).  
 GN GPSA OR DR2621.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
 OX Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D., Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L., Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M., Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C., Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D., Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C., Fraser C.M.;  
 RA "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.";  
 RT Science 286:1571-1577(1999).  
 RL CC -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(P)H -> glycerone phosphate + NAD(P)H.  
 CC -!- PATHWAY: De novo phospholipid biosynthesis; glycerol-3 phosphate formation.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE DEHYDROGENASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AE002091; AAF12158.1;  
 DR TIGR; DR2621;  
 DR InterPro; IPR001652; NAD\_Gly3P\_dh.  
 DR Pfam; PF01210; NAD\_Gly3P\_dh; 1.  
 DR PRINTS; PR00077; GPDHGRNASE.  
 DR ProDom; PD001649; NAD\_Gly3P\_dh; 1.  
 DR PROSITE; PS00957; NAD\_G3PDH; 1.  
 KW Phospholipid biosynthesis; Oxidoreductase; NAD; Complete proteome.  
 SQ SEQUENCE 328 AA; 33597 MW; 46D909D7D14170BE CRC64;  
 Query Match 6.7%; Score 88.5; DB 1; Length 328;





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.3688 Seconds  
(without alignments)  
2526.317 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPFLAVDIGNTTVLGL.....DETLRLGLVLMASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1324	100.0	262	16	Q9RX54
2	437.5	33.0	254	16	Q9KGH5
3	414.5	31.3	255	16	Q8R7M2
4	412.5	31.2	258	2	Q9F985
5	409.5	30.9	265	16	Q9X8N6
6	404	30.5	272	16	O06282
7	401	30.3	274	16	Q9CD56
8	396.5	29.9	261	16	Q9A6Z1
9	385.5	29.1	259	16	Q8XHL5
10	371.5	28.1	273	16	Q97EB4
11	363.5	27.5	259	16	O8YAC5
12	356.5	26.9	259	16	O92F54
13	344.5	26.0	256	16	Q8RFE4
14	271	20.5	212	2	O32514
15	260	19.6	246	16	Q9WZY5
16	248	18.7	273	16	O83446

17	159	12.0	262	16	O51477
18	152.5	11.5	276	16	Q8YQD7
19	139.5	10.5	295	16	Q8Y2M4
20	133.5	10.1	229	16	O67753
21	128.5	9.7	242	16	O9PC14
22	123.5	9.3	257	16	P74045
23	107	8.1	438	17	Q8TS02
24	104.5	7.9	382	16	O8XDP0
25	104.5	7.9	400	16	O86549
26	104	7.9	505	3	Q9Y799
27	103.5	7.8	255	16	O9L1E6
28	101.3	7.7	444	16	O9A5G5
29	101	7.6	435	16	Q92L66
30	101	7.6	636	2	O53818
31	101	7.6	636	16	O8Q039
32	101	7.6	1441	16	O9CFL1
33	100.5	7.6	592	16	O9JW17
34	100	7.6	615	2	O9X618
35	99.5	7.5	544	17	O9Y8G2
36	99.5	7.5	1363	2	O9F5K8
37	98.5	7.4	302	16	P96939
38	98.5	7.4	477	16	O9L211
39	97	7.3	323	16	O99YU9
40	96.5	7.3	515	16	O8YDQ5
41	96.5	7.3	592	16	O9JXF1
42	95	7.2	297	2	O93FF7
43	95	7.2	2065	2	O93DC7
44	95	7.2	2691	16	O8XP11
45	95	7.2	11096	2	O9L4W3

## ALIGNMENTS

## RESULT 1

Q9RX54	Q9RX54	PRELIMINARY;	PRT;	262 AA.
ID	Q9RX54;			
AC	Q9RX54;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Hypothetical protein DR0461.			
GN	DR0461.			
OS	Deinococcus radiodurans.			
OC	Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;			
OC	Deinococcaceae; Deinococcus.			
OX	NCBI_TaxID=1299;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R1;			
RX	MEDLINE=20036896; PubMed=10567266;			
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,			
RA	Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,			
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,			
RA	Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,			
RA	Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,			
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,			
RA	Fraser C.M.;			
RT	"Genome sequence of the radioresistant bacterium Deinococcus			
RT	radiodurans R1."			
RL	Science 286:1571-1577(1999).			
DR	EMBL; AE001905; AAF10040.1; -			
DR	TIGR; DR0461.			
DR	InterPro; IPR004619; Baf.			
DR	InterPro; IPR001220; Lectin_legB.			
DR	Pfam; PF03309; Bvg_acc_factor; 1.			
DR	TIGRFAMs; TIGR00671; baf; 1.			
DR	PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 262 AA; 27839 MW; 965EAD2F78785A0 CRC64;			
Query Match	100.0%; Score 1324; DB 16; Length 262;			
Best Local Similarity	100.0%; Pred. No. 2.5e-95;			

O51477	borrella bu
O8YQD7	anabaena sp
O8Y2M4	raistonia s
O67753	aquifex aeo
O9PC14	xylella fas
P74045	synechocyst
Q8TS02	methanosarc
O8XDP0	raistonia s
O86549	streptomyce
Q9Y799	candida par
O9L1E6	streptomyce
O9A5G5	caulobacter
Q92L66	rhizobium m
Q53818	streptomyce
O8Q039	streptomyce
O9CFL1	lactococcus
O9JW17	neisseria m
O9X618	streptomyce
Q9Y8G2	aeropyrum p
Q9F5K8	rhodobacter
P96939	mycobacteri
O9L211	streptomyce
O99YU9	streptococ
O8YDQ5	brucella me
O9JXF1	neisseria m
O93FF7	acineto
O93DC7	yersinia en
O8XP11	raistonia s
O9L4W3	streptomyce

Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAPPLAVDIGNTTVGLADASGALHTWRTNREMLPDDLALQHLGFTLAGAPIP 60  
 DB 1 MPAPPLAVDIGNTTVGLADASGALHTWRTNREMLPDDLALQHLGFTLAGAPIP 60

QY 61 RAAVLSSVAPPVGENVALAKRHFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120  
 DB 61 RAAVLSSVAPPVGENVALAKRHFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 120

QY 121 EKYLGGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAALPRITLQA 180  
 DB 121 EKYLGGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAALPRITLQA 180

QY 181 PETAICKNTVHALQSLVFCYAEWVDGLLRIRAEPLGAEVAVATGFSRTVQICQEID 240  
 DB 181 PETAICKNTVHALQSLVFCYAEWVDGLLRIRAEPLGAEVAVATGFSRTVQICQEID 240

QY 241 YDETLTLRGLVLSRSEVR 262  
 DB 241 YDETLTLRGLVLSRSEVR 262

RESULT 2  
 Q9KGH5 PRELIMINARY; PRT; 254 AA.

AC Q9KGH5; 01-OCT-2000 (Tremblrel..15, Created)  
 DT 01-OCT-2000 (Tremblrel..15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel..21, Last annotation update)  
 DE Hypothetical protein BH0086.  
 GN BH0086.

OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AP001507; BAB03805.1; --  
 DR InterPro; IPR004619; Baf.  
 PF Pfam; PF03309; Bvg\_acc\_factor; 1.  
 TG TIGRFAMs; TIGR00671; baf; 1.  
 AA Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 33.0%; Score 437.5; DB 16; Length 254;  
 Best Local Similarity 35.7%; Pred. No. 2.4e-26;  
 Matches 89; Conservative 56; Mismatches 97; Indels 7; Gaps 4;

QY 6 LLAVDIGNTTVGLADASGALHTWRTNREMLPDDLALQHLGFTLAGAPIP--RAA 63  
 DB 2 ILVIDGNTNTVLGVYQ--DETLVHWRLATSRQKTEDEYAMTVRSFDHAGLQFQIDGI 60

QY 64 VLSSVAPPVGENVALAKRHFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 123  
 DB 61 VISSVPPPMFSLQOMCKKYFHTPMIIGPKITGLNRYDNPKEVGADRLVNAVAIEL 120

QY 124 LGGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAALPRITLQA 182  
 DB 121 YG---YPAIVDFGTATCLINEKQYAGGVAPGIMISTEALYHRASKLPRIEAKPK 177

QY 183 TAIGKNTVHALQSLVFCYAEWVDGLLRIRAEPLGAEVAVATGFSRTVQICQEID 242  
 DB 178 QVVGNTIDMSQSGIFGYGVSDGVVKKKAQAESEPKVIATGGLAKLIGTESIDVI 237

QY 243 DETILTRGL 251  
 DB 238 DSFLTLKGL 246

RESULT 3  
 Q8R7M2 PRELIMINARY; PRT; 255 AA.

AC Q8R7M2; 01-JUN-2002 (Tremblrel..21, Created)  
 DT 01-JUN-2002 (Tremblrel..21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel..21, Last annotation update)  
 DE Putative transcriptional regulator, homologs of Bvg accessory  
 DE factor.  
 GN TFE2381.  
 OS Thermoanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
 OX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB4T / JCM11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome."  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL; AE013180; AAM25520.1; --  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 31.3%; Score 414.5; DB 16; Length 255;  
 Best Local Similarity 35.8%; Pred. No. 1.5e-24;  
 Matches 92; Conservative 48; Mismatches 112; Indels 5; Gaps 3;

QY 6 LLAVDIGNTTVGLADASGALHTWRTNREMLPDDLALQHLGFTLAGAPIPRA--A 63  
 DB 2 LLAFDVGNTNVMGVFKG-KLLHSFRISTDKNTYDEYGLVQLNGLYNGISLTDVV 60

QY 64 VLSSVAPPVGENVALAKRHFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFGA 123  
 DB 61 IISVVPPLMTLQVMSLRYFKPIVVGPKITGINKIKYDNPKEVGADRLVNAVAIEL 120

QY 124 LGGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSADALFARAALPRITLQA 183  
 DB 121 YGG--PVIVIDFGTATTCALSEKEGYLGGIAPGLMSADALFQRTAKLPKIDLT 178

QY 184 AIGKNTVHALQSLVFCYAEWVDGLLRIRAEPLGAEVAVATGFSRTVQICQEID 243  
 DB 179 VINRNTVASMQSGIYGHVGMVDYIVTRMGEFAPSAYVATGGFANNIAESKTDITVN 238

QY 244 ETILTRGLVLSRSE 260  
 DB 239 EMUTLEGLRIIYERNE 255

RESULT 4  
 Q9F985 PRELIMINARY; PRT; 258 AA.

AC Q9F985; 01-MAR-2001 (Tremblrel..16, Created)  
 DT 01-MAR-2001 (Tremblrel..16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel..21, Last annotation update)  
 DE Putative 32 kDa replication protein.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Geobacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V;  
 RA Vasquez C., Pichuanes S., Saavedra C.;



RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF198621; AAG28531.1; -;  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs: TIGR00671; baf; 1.  
 SQ SEQUENCE 258 AA; 28101 MW; 5078550695095855 CRC64;

Query Match 31.2%; Score 412.5; DB 2; Length 258;  
 Best Local Similarity 35.5%; Pred. No. 2.2e-24;  
 Matches 93; Conservative 48; Mismatches 96; Indels 25; Gaps 6;

QY 6 LLAVDIGNTTVLGLADASGALHTWRIPNREMLPDDLALQLHGLFTLAGAPIP--RAA 63  
 DB 2 IFVLDIGNTTVLGVYDGD--ELKHHWRIETSRKTEDEYGMTKALLNHVGLQFSDIDGI 60  
 QY 64 VLSSVAPPVGENYALAKR----HFMDFAVSAENLPDVTVELDTPGSGVADRLCN-- 116  
 DB 61 IISVVVPI-----MFALERMCLAYFHKPIIVGPGIKTGLNIKYDNPREGVADRIYNVA 116  
 DB 117 ---LFGAEKYLGLDYAVVDFGTSTNFDVVGRRRFLGGLATGAQVSADALFARA 173  
 DB 117 GIHLGSP-----LIIVDFGTATTTCYINEHKQYMGGAIAPIGIMISTEALFARA 168  
 QY 174 PRITLQAPETAICKNTVHALQSLVFGYAEWVDGLLRIRAEPLGAEVAVATGGFSRTVQ 233  
 DB 169 PRIETIARPDIIIGKNTVSAMQAGILYGVQVEGIVSRMKAKSPVPPKVIATGGLASLIA 228  
 QY 234 GICOEDIDYDETLTLRGLVELW 255  
 DB 229 SESNVIDIVDPFLTLGLKILY 250

# RESULT 5

Q9X8N6 PRELIMINARY; PRT; 265 AA.  
 AC Q9X8N6;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein SCO3380.  
 GN SCO3380 OR SCE94.31C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 AL Oliver K., Harris D.;  
 AL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapait D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RA "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 RL EMBL: AL049628; CAB40880.1; -;  
 DR InterPro: IPR004619; Baf.  
 DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
 DR TIGRFAMs: TIGR00671; baf; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 30.9%; Score 409.5; DB 16; Length 265;  
 Best Local Similarity 37.9%; Pred. No. 3.9e-24;  
 Matches 97; Conservative 44; Mismatches 102; Indels 13; Gaps 4;

QY 6 LLAVDIGNTTVLGLADASGALHTWRIPNREMLPDDLALQLHGLFTL-----AGAP 58  
 DB 2 LLTIDVGNTHVVLGDFGEDIH-WRISTDSRTADELAVLLQGLMGHPPLGDELGDG 60  
 QY 59 IPRAAVLSSVAPPVGENYALAKRHFMDFAVSAENLPDVTVELDTPGSGVADRLCNLF 118  
 DB 61 IDGTAICATVPSVLHELREVTTRYGDPVAVLVEPGVKTGPILTDHPKEVGADRIINAV 120  
 QY 119 GAEKYLGLDYAVVDFGTSTNFDVVGRRRFLGGLATGAQVSADALFARA 178  
 DB 121 AAVELYGG--PAIVDFGTATTDAVSARGEYIGVIAPIGIEISVEALGVGAOLRKIEV 178  
 QY 179 QAPETAICKNTVHALQSLVFGYAEWVDGLLRIRAEI---PGEAVAVATGGFSRTVQGI 235  
 DB 179 ARPSRVIGKNTVEAMQSGIVYGFAGQVDGVVNRWARELADDDVTVIATGGLAPWLGE 238  
 QY 236 COEIDYDETLTLRGL 251  
 DB 239 SSVIDEHEPWLTLMLG 254

# RESULT 6

O06282 PRELIMINARY; PRT; 272 AA.  
 ID O06282;  
 AC O06282;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative,  
 DE Baf family).  
 GN RV3609C OR MTCY07H7B.22 OR MT3706.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Brown T., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-CDC 1551 / OSHKOSH;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,





[illegible]

QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAE-NLPDVTVELDTPGSGVADRLCNLFGE-E 121  
DB 62 LISSVVPNIITTFQFPARKYFKVEATVDEKKLPFTFAKGINVTGFGADRIIDTEAMQ 121  
QY 122 KYLGGLDYAVVDFGTSTNFDVVGRRRLFGGLTATGAQVSADALFARAARAKLPRLTQAP 181  
DB 122 KYPD--KNLVDFDFTATTYDLVLRG-KVYGILPGIDMSINALYGNKALPRVRFETTP 178  
QY 182 ETAIGNKNTVHALQSLGVFGYAEVMDGLLRRAELRRAELPGEAVAVATGFSRTVQICQIEDY 241  
DB 179 SSVLGTDWKQQAQAFYAGQIKHIIKINEELNEEFVLTATGGILGKILSAEIDEIDE 238  
QY 242 YDETLRLGLVELM 255  
DB 239 YDANLSLKGLYTLY 252

RESULT 14  
Q32514 PRELIMINARY; PRT; 212 AA.  
O32514;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Similar to Bacillus subtilis.  
OS Desulfovibrio vulgaris (strain Miyazaki).  
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;  
OC Desulfovibrio.  
OX NCBI\_TaxID=883;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MIYAZAKI;  
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,  
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;  
RT "Sequence analyses of two ferredoxin genes and their flanking regions  
RT from Desulfovibrio vulgaris (Miyazaki F.).";  
RL J. Biochem. Mol. Biol. Biophys. 2:147-154 (1998).  
RL EMBL; AB005550; BAA21476.1;  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

Query Match 20.5%; Score 271; DB 2; Length 212;  
Best Local Similarity 36.5%; Pred. No. 1.7e-13;  
Matches 77; Conservative 31; Mismatches 89; Indels 14; Gaps 7;

QY 7 LAVDIGNTTTVGLADASGALTHTWRTNREMLPDDLLALQLHGLFTLAG-APIP-RAAV 64  
DB 6 LLEDIGNTNVKGIAVETAVLT-SYVLPDTPGQTTDSIGLRLLLEVLRHAGLGPADVGACV 64  
QY 65 LSSVAPPVGENYALAKRHFMDAFVSAE-NLPDVTVELDTPGSGVADRLCNLFGE 120  
DB 65 ASSVPGVNPILRRACERYLYRLLFPAG----DIALPDRYERPAEVGADRLVAAYAA 120  
QY 121 EKYLGGLDYAVVDFGTSTNFDVVGRRRLFGGLTATGAQVSADALFARAARAKLPRLTQAP 180  
DB 121 RLRYPGPRSLVSDFGTATFDVCV-EGGAYILGLICPGVLSSAGALSSRTAKLPRLISLEV 179  
QY 181 PETA--IGKNTVHALQSLGVFGYAEVMDGLLRRAELRRAELPGEAVAVATGFSRTVQICQIEDY 241  
DB 180 EEDSPVIGRSTTSLNHGFIIFGFAAMTEGVL 210

RESULT 15  
Q9WZ15 PRELIMINARY; PRT; 246 AA.  
AC Q9WZ15;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein TM0883.  
GN TM0883.  
OS Thermotoga maritima.

OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-KS8 / DSM 3109;  
RX MEDLINE-99287316; PubMed-10360571;  
RA NELSON K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329 (1999).  
DR EMBL; AE001754; AAD35964.1;  
DR TIGR; TM0883;  
DR InterPro: IPR004619; Baf.  
DR Pfam: PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMS; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;  
Query Match 19.6%; Score 260; DB 16; Length 246;  
Best Local Similarity 30.6%; Pred. No. 1.5e-12;  
Matches 75; Conservative 48; Mismatches 114; Indels 8; Gaps 7;  
QY 7 LAVDIGNTTTVGLADASGALTHTWRTNREMLPDDLLALQLHGLFTLAGAPIPRAAVLS 66  
DB 3 LLDVGNTHSVFSITE-DGKTFRWRLSTGVFQTEDELFSLHPLLDGAMREIKGIV-A 60  
QY 67 SVAPPVGENYALAKRHFMDAFVSAE-NLPDVTVELDTPGSGVADRLCNLFGEAKYLG 126  
DB 61 SVPTQNTVIERFSQKIFHISPIWVAKN-GCVKNVKNPSEVGADRVANVAFVKEYG- 118  
QY 127 LDYAVVDFGTSTNFDVVGRRRLFGGLTATGAQVSADALFARAARAKLPRLTQAPETAIG 186  
DB 119 -KNGIIIDMGTTATVDLVVNG-SYEGGAILPFGFMVHSLFRGTAKLPLEVEKPADFVVG 176  
QY 187 KNTVHALQSLGVFGYAEVMDGLLRRAELRRAELPGEAVAVATGFSRTVQICQIEDYDETL 246  
DB 177 KTEENIRLGVVNGSVYALEGIIGRIK-EVYGLDPLVLTGGOSKIVKDMIKH-EIFDEDL 234  
QY 247 TLRLG 251  
DB 235 TIKGV 239

Search completed: June 24, 2003, 21:59:22  
Job time : 22.3688 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.899 Seconds  
(without alignments)  
1297.879 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPFLAVDIGNTTTLGL.....DETILRLGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

---searched: 908470 seqs, 133250620 residues. 908470  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002: \*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: \*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1324	100.0	262	23 AAU911155	Deinococcus radiop
2	437.5	33.0	254	23 AAU911171	Pantothenate kins
3	428	32.3	258	22 AAU01243	B. subtilis novel
4	428	32.3	258	23 AAU911149	Bacillus subtilis
5	415.5	31.4	258	23 AAU911172	Pantothenate kins
6	414.5	31.3	255	23 AAU911154	Geobacter sulfure
7	409.5	30.9	265	23 AAU911151	Streptomyces coeli
8	405.5	30.6	262	23 AAU911170	Pantothenate kins
9	404	30.5	272	22 AAG81225	Mycobacterium tube
10	404	30.5	272	23 AAU911152	Mycobacterium tube

11	396.5	29.9	260	23 AAU911173	Pantothenate kins
12	378.5	28.6	233	23 AAU911163	Pantothenate kins
13	377	28.5	256	23 AAU911175	Pantothenate kins
14	363.5	27.5	259	23 ABB47661	Listeria monocytog
15	347.5	26.2	258	23 AAU911153	Rhodobacter capsul
16	337	25.5	250	23 AAU911150	Clostridium acetob
17	333.5	25.2	219	23 AAU911176	Pantothenate kins
18	271	20.5	212	23 AAU911177	Pantothenate kins
19	260	19.6	246	23 AAU911156	Thermotoga maritim
20	253.5	19.1	257	23 AAU911174	Pantothenate kins
21	248	18.7	273	23 AAU911157	Treponema pallidum
22	215	16.2	241	23 AAU911179	Pantothenate kins
23	159	12.0	262	23 AAU911158	Borrelia burgdorfe
24	133.5	10.1	229	23 AAU911159	Aquifex aeolicus p
25	131	9.9	244	23 AAU911168	Pantothenate kins
26	128.5	9.7	242	23 AAU911180	Pantothenate kins
27	123.5	9.3	257	23 AAU911160	Synechocystis pant
28	107	8.1	457	22 AAU56858	Propionibacterium
29	104	7.9	505	19 AAU70516	Candida parapsilos
30	104	7.9	505	22 AAG64143	Candida parapsilos
31	103.5	7.8	267	23 AAU911162	Bordetella pertussis
32	101.5	7.7	455	20 AAY38617	Neisseria gonorrhoe
33	101.5	7.7	455	21 AAY74908	Neisseria gonorrhoe
34	101.5	7.7	460	23 AAU911167	Pantothenate kins
35	101.5	7.7	592	20 AAY38618	Neisseria gonorrhoe
36	101.5	7.7	592	21 AAY74911	Neisseria gonorrhoe
37	101	7.6	1440	23 ABB54801	Lactococcus lactis
38	100.5	7.6	455	21 AAY74910	Neisseria meningit
39	100.5	7.6	592	20 AAY38616	Neisseria meningit
40	100.5	7.6	592	21 AAY74913	Neisseria meningit
41	100.5	7.6	592	23 AAU911166	Pantothenate kins
42	99	7.5	984	23 AAU19794	Human Hairless pro
43	99	7.5	1189	20 AAU15217	Human Hairless wil
44	99	7.5	1189	23 AAU19796	Human Hairless pro
45	98	7.4	249	23 AAU911182	Pantothenate kins

ALIGNMENTS

RESULT 1  
AAU911155  
ID AAU91155 standard; Protein; 262 AA.  
XX AC  
XX AAU91155;  
XX 05-JUN-2002 (first entry)  
XX DT  
XX DE Deinococcus radiopugnans pantothenate kinase Coax.  
XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
XX KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX OS Deinococcus radiopugnans.  
XX XX  
XX WO200216601-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26531.  
XX PR 24-AUG-2000; 2000US-227860P.  
XX PR 20-MAR-2001; 2001US-0813453.  
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
XX PA Yocum RR, Patterson TA;  
XX PI WPI; 2002-269358/31.  
XX DR Identifying potential antibiotic or antimicrobial agent, comprises  
XX PT contacting composition comprising pantothenate kinase (Coax) protein  
XX PT with test compound and identifying inhibitor of the Coax protein -





DR WPI; 2001-218644/22.  
DR N-PSDB; AAS00984.

XX New recombinant microorganism which overexpress a *Bacillus subtilis*  
PT pantothenate biosynthetic enzyme, useful for the high yield production  
PT of panto-compounds such as pantothenate and pantoate.  
XX

PS Example 14; Fig 23; 292pp; English.

XX The sequence represents a novel *B. subtilis* pantothenate kinase (encoded  
CC by gene cox), an enzyme of the pantothenate biosynthetic pathway.  
CC Pantothenate, also known as vitamin B5, is used as a nutritional  
CC supplement in mammals and humans. The invention concerns methods of  
CC producing recombinant microorganisms overexpressing at least one *Bacillus*  
CC *subtilis* pantothenate biosynthetic enzyme. The microorganisms and methods  
CC of producing them are useful for producing a panto-compound such as  
CC pantothenate or pantoate, which is a nutritional requirement for  
CC livestock and humans. The methods are also useful for the identification  
CC of pantothenate kinase modulators. Panto-compounds are produced at a  
CC significantly higher yield than prior art methods and can be produced  
CC independent of the need to feed precursors which decreases expense.

XX Sequence 258 AA;

Query Match 32.3%; Score 428; DB 22; Length 258;  
Best Local Similarity 37.4%; Pred. No. 1.4e-36;  
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

QY 6 LLAVDIGNTTVLGLADASGALHTWRTNREMLPDDLLALQLHGLFTLAGAPIPR--AA 63  
DB 2 LLVIDGNTNTVLGVYH-DKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60  
QY 64 VLSSVAPPVGENYALAKR---HFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFG 119  
DB 61 IISVVVPI---MFALERMCTKYFHIEPQIVGPGMKTGLNKNYDNPKEVGADRIVNAVA 116  
QY 120 AKYLGGLDYAVVVDGTSNFDVVGRRFLGGLATGAQVSDADALFARAALPRITLQ 179  
DB 117 AIHLYG--NPLIVVDGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIET 174  
QY 180 APETAIGKNTVHALQSLGVFAEMVDGLLRRAELPGEAVAVATGSGFSRTVQICQEI 239  
DB 175 RPDNIIGNTVSAMQSGILFGYGVQGVGIVKRMKQAKQDLKVIATGGLAPLIANESDCI 234  
QY 240 DYDETLTLRGLVELWASREV 261  
DB 235 DIVDPFLLKGL-ELIYERNRV 255

RESULT 4  
AAU91149  
ID AAU91149 standard; Protein; 258 AA.  
XX  
AC AAU91149;

XX 05-JUN-2002 (first entry)

XX *Bacillus subtilis* pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX *Bacillus subtilis*.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;

XX WPI; 2002-269358/31.

DR N-PSDB; ABK54168.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein  
XX  
XX Claim 10; Page 67-68; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein.  
CC described in the invention.

XX Sequence 258 AA;

Query Match 32.3%; Score 428; DB 23; Length 258;  
Best Local Similarity 37.4%; Pred. No. 1.4e-36;  
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

QY 6 LLAVDIGNTTVLGLADASGALHTWRTNREMLPDDLLALQLHGLFTLAGAPIPR--AA 63  
DB 2 LLVIDGNTNTVLGVYH-DKLEYHWRIETSRHKTDEFGMLRSLFDHSGLMFEQIDGI 60  
QY 64 VLSSVAPPVGENYALAKR---HFMIDAFVSAENLPDVTVELDTPGSGADRLCNLFG 119  
DB 61 IISVVVPI---MFALERMCTKYFHIEPQIVGPGMKTGLNKNYDNPKEVGADRIVNAVA 116  
QY 120 AKYLGGLDYAVVVDGTSNFDVVGRRFLGGLATGAQVSDADALFARAALPRITLQ 179  
DB 117 AIHLYG--NPLIVVDGTTATTCYIDENKQYMGGAIPGTTISTEALYSRAAKLPRIET 174  
QY 180 APETAIGKNTVHALQSLGVFAEMVDGLLRRAELPGEAVAVATGSGFSRTVQICQEI 239  
DB 175 RPDNIIGNTVSAMQSGILFGYGVQGVGIVKRMKQAKQDLKVIATGGLAPLIANESDCI 234  
QY 240 DYDETLTLRGLVELWASREV 261  
DB 235 DIVDPFLLKGL-ELIYERNRV 255

RESULT 5  
AAU91172  
ID AAU91172 standard; Protein; 258 AA.  
XX  
AC AAU91172;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #10.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
XX *Bacillus stearothermophilus*.

XX WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX

```

PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX N-PSDB; ABK54193.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 10; Page 101-102; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 258 AA;
Query Match 31.4%; Score 415.5; DB 23; Length 258;
Best Local Similarity 35.2%; Pred. No. 2.9e-35;
Matches 94; Conservative 49; Mismatches 99; Indels 25; Gaps 6;
QY 6 LLAVDIGNTTTLVLGADASGALHTWIRTNREMLPDDLALQLHGLFTLAGAPIP--RAA 63
DB 2 IFVLVDGNTNVLGVYDGD-ELKHHWRIETSRKTEDEYGMKALLNHHVGLQFSDIRGI 60
QY 64 VLSSVAPPVGENYALAKR---HEMIDAFVSAENLPDVTVELDTPGSGVADRCLN---116
DB 61 IISVVPPVPI---MFALERMKLKFHFKPLVIGPGIKTGDKIDKYNPREVGADRIVNAVA 116
QY 117 ---LFGAEKYLGLLDYAVVVDFTSTNFDVYGRGRRLGGIILATGAQVSADALFAAAKL 173
DB 117 GIHLVGSF-----LIIVDFGATTCYINHEKHQYMGGAIPGIMISTEALFAAAKL 168
QY 174 PRITLQAPETAIGNKTVHALQSLGVFGYAEWVDGLLRIRRAELPGEAVAVATGGFSRTVQ 233
DB 169 PRIEARPDDIIGNKTVSAMQAGILYGVGVGEGIVSRMKAKSKIPPRVIATGGLAPLIA 228
234 GICQEDYDDETLLRLGLVELWASRSE 260
229 SESDIIVVDVDFLTGLTKLYEKNT 255

```

Geobacter sulfurreducens pantothenate kinase Coax.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Geobacter sulfurreducens.

WO200216601-A2.

28-FEB-2002.

24-AUG-2001; 2001WO-US26531.

```

XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein
XX
PS Claim 10; Page 72-73; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 255 AA;
Query Match 31.3%; Score 414.5; DB 23; Length 255;
Best Local Similarity 36.8%; Pred. No. 3.7e-35;
Matches 96; Conservative 50; Mismatches 102; Indels 13; Gaps 6;
QY 6 LLAVDIGNTTTLVLGADASGALHTWIRTNREMLPDDLALQLHGLFTLAGAPIP--RAA 63
DB 2 LLVIDVGNNTNVLGVYDGE-RLVRDWRVSTDRARTDEVGILLNELFRLAGLGLQIRAV 60
QY 64 VLSSVAPP---VGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRCLN-LFG 119
DB 61 IISVVPPVPLTGLYERLSLG---YFGMRPLVGVGPGIKTGMPIDYNDPREVGADRIVNAV 117
QY 120 AEKYLGLLDYAVVVDFTSTNFDVYGRGRRLGGIILATGAQVSADALFAAAKLPRITLQ 179
DB 118 YEKYTSLS---LIIVDFGATTCYINHEKHQYMGGAIPGLVISTEALFORASKLPVDII 174
QY 180 APETAIGNKTVHALQSLGVFGYAEWVDGLLRIRRAELPGEAVAVATGGFSRTVQICQEI 239
DB 175 RPSALIARTVNSMQAGIYGVGVGLVDEIVTRMKAESKDAPRVATGGLASLIAPESKTI 234
240 DYYDETLLRLGLVELWASRSE 260
235 EAVEEYLTLEGRLIYERNRE 255

```

Streptomyces coelicolor pantothenate kinase Coax.

Pantothenate kinase; Coax; antibiotic; antimicrobial;

pantothenate kinase modulator; coenzyme A; bactericidal compound.

Streptomyces coelicolor.

WO200216601-A2.

28-FEB-2002.

```

RESULT 6
AAU91154
ID AAU91154 standard; Protein; 255 AA.
XX
AC AAU91154;
XX
DT 05-JUN-2002 (first entry)
XX
DE Geobacter sulfurreducens pantothenate kinase Coax.
XX
KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Geobacter sulfurreducens.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.

```

```

PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
PS Claim 10; Page 69-70; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 265 AA;

Query Match 30.9%; Score 409.5; DB 23; Length 265;
Best Local Similarity 37.9%; Pred. No. 1.3e-34;
Matches 97; Conservative 44; Mismatches 102; Indels 13; Gaps 4;

QY 6 LLAVDIGNTTTLVLGADASGALHTWRITNREMLPDDLALQLHGLFTL-----AGAP 58
DB ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 LITIDVGNTHVTLGDFGEDIYEH-WRISTDSRRTADELAVLVLQGLMGHPLLGDELGDG 60

QY 59 IPRAVLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLF 118
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
61 IDGIAICATVPVSLHELREVTTRYGYDPAVLVEPGVKTVGVPILTDPKVEVGADRIINAV 120

QY 119 GAEKYLGGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAALPRITL 178
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 AAVELYGG--PAIVVDFTGATTDFDAVSARGEYIGGVIAPIGIEISVEALGVGAQLRKIEV 178

QY 179 CAPETAIGNKNTVHALQSLGVFGYAEWVDGLLRIRAEEL---PGEAVAVATGFSRTVQGI 235
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
179 APRSRVIGNKNTVEAMQSGILYGVQGVGVNRMARELADDDPDDVTVIATGGLAPMVLGE 238

QY 236 COEIDYDETTLRGL 251
DB ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
239 SSVIDEHEPWLTMGL 254

RESULT 8
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX
AC AAU91170;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (Coax) #8.
XX
XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
XX Bacillus anthracis.
XX
XX WO200216601-A2.
XX
XX 28-FEB-2002.

```

```

XX 24-AUG-2001; 2001WO-US26531.
XX
XX 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54191.
XX
XX Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (Coax) protein
PT with test compound and identifying inhibitor of the Coax protein -
XX
XX Claim 8; Page 98-99; 128pp; English.
XX
XX The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (Coax) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the Coax protein, an
CC essential enzyme for the production of coenzyme A. Coax protein is a
CC valuable target for identifying bactericidal compounds. Coax modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (Coax) protein
CC described in the invention.
XX
SQ Sequence 262 AA;

Query Match 30.6%; Score 405.5; DB 23; Length 262;
Best Local Similarity 31.7%; Pred. No. 3.4e-34;
Matches 84; Conservative 60; Mismatches 104; Indels 17; Gaps 4;

QY 6 LLAVDIGNTTTLVLGADASGALHTWRITNREMLPDDLALQLHGLFTLAGAPTP--RAA 63
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2 IFVLDVGNTHVTLGDFGEDIYEH-WRISTDSRRTADELAVLVLQGLMGHPLLGDELGDG 60

QY 64 VLSSVAPPVGENYALALKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCN-----L 117
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
61 IVSVVPPINFALERCEKYEKIKPLVVGPIKTLNKKYENPREVGADRIINAVAGIHL 120

QY 118 FGAEKYLGLDYAVVVDFTSTNFDVVGRRRFLGGILATGAQVSADALFARAALPRIT 177
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
121 YGSP-----LIIVDFGTATTTCYINEEKHYMGVITPGIMISAEALYSRAAKLPRIE 172

QY 178 LQAPETAIGNKNTVHALQSLGVFGYAEWVDGLLRIRAEELPGEAVAVATGFSRTVQGI 237
DB |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
173 ITRPSSVVGKNTVSAHQSGILYGVQGVGVNRMARELADDDPDDVTVIATGGLAPMVLGE 238

QY 238 EIDYDETTLRGLVLEWASRSEVR 262
DB ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
233 VIDVDFPLKGLMLYERNALQ 257

RESULT 9
AAG81225
ID AAG81225 standard; Protein; 272 AA.
XX
AC AAG81225;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 276.
XX
XX Drug target; growth; organism viability; characterisation.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200135317-A1.
XX

```

PD 17-MAY-2001.  
 XX 13-NOV-2000; 2000WO-US31152.  
 XX 12-NOV-1999; 99US-0165086.  
 PR 12-NOV-1999; 99US-0165124.  
 PR 01-FEB-2000; 2000US-0179531.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Eisenberg D, Rotstein SH, Marcotte EM;  
 PI WPI: 2001-329193/34.  
 XX N-PSDB; AAH52076.  
 DR  
 XX  
 PT Identifying nucleotide or polypeptide sequence for use as drug target,  
 PT involves providing algorithm that analyzes a functional relationship  
 PT between nucleotide or polypeptide sequences, and comparing the  
 PT sequences  
 XX  
 Disclosure: Page 188; 207pp; English.  
 CC This invention relates to a method for identifying a nucleotide or  
 CC polypeptide sequence that may be a drug target, or essential for growth  
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
 CC tuberculosis proteins which are potential drug targets. The DNA and  
 CC protein sequences are used to illustrate the method of the invention. The  
 CC method involves providing an unknown nucleotide or polypeptide sequences,  
 CC and comparing it to a number of sequences along with at least one  
 CC algorithm capable of analyzing a functional relationship between one  
 CC nucleotide and polypeptide sequences. The method is useful for  
 CC characterizing the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism.  
 XX  
 XX Sequence 272 AA;  
 SQ  
 Query Match 30.5%; Score 404; DB 22; Length 272;  
 Best Local Similarity 38.3%; Pred. No. 5.1e-34;  
 Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;  
 QY 6 LLAVDIGNTTTLGLADASGALHT-----WRIRNREMLPDDLAQLHGLFTLAGAPIP 60  
 DB 2 LLAVDRNTHTVVGL--LSGKHEAKVQVQWRIRTESEVTADELALTDGLIGDSERLT 59  
 QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTPGSVG 110  
 DB 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIPEGVRTGIPLLVDNPKVEG 109  
 QY 111 ADRLCNLFGAEKYLGGLDYAVVDFGTSTNFDVVGRRRFLGILATGAQVSADALFARA 170  
 DB 110 ADRIVNCCLAA--YDRFRKAAIVDFGSSICVDVVSAGKEFLGAGTAPGVQVSSDAAARS 167  
 QY 171 AKLPRTITQAPETAIGKNTVHALQSLGVFGAEVMDGLLRIRRAELPG-----EAVAVAT 225  
 DB 168 AALRRVELARPSVVGKNTVECMQAGVFGAGLVGLVGRIDREYVSGFSVDHDAIVAT 227  
 QY 226 GGFSTVQICQEDYDDETTLRGLVLMASRSEVR 262  
 DB 228 GHTAPLLPELHTVDHQHLLTQGLRLVFERNLEVQ 264  
 RESULT 10  
 AAU91152  
 ID AAU91152 standard; Protein; 272 AA.  
 AC AAU91152;  
 XX  
 XX 05-JUN-2002 (first entry)  
 DT Mycobacterium tuberculosis pantothenate kinase Coax.  
 DE  
 XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO200216601-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 XX 24-AUG-2001; 2001WO-US26531.  
 PF  
 XX 24-AUG-2000; 2000US-227860P.  
 PR  
 PR 20-MAR-2001; 2001US-0813453.  
 XX  
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 XX  
 PI Yocum RR, Patterson TA;  
 XX  
 DR WPI: 2002-269358/31.  
 XX  
 PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein  
 XX  
 PS Claim 10; Page 70-71; 128pp; English.  
 XX  
 CC The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX  
 XX Sequence 272 AA;  
 SQ  
 Query Match 30.5%; Score 404; DB 23; Length 272;  
 Best Local Similarity 38.3%; Pred. No. 5.1e-34;  
 Matches 106; Conservative 42; Mismatches 95; Indels 34; Gaps 7;  
 QY 6 LLAVDIGNTTTLGLADASGALHT-----WRIRNREMLPDDLAQLHGLFTLAGAPIP 60  
 DB 2 LLAVDRNTHTVVGL--LSGKHEAKVQVQWRIRTESEVTADELALTDGLIGDSERLT 59  
 QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDTPGSVG 110  
 DB 60 GTAALSTV-PSVLHEVRIMLDQYW-----PSVPHVLIPEGVRTGIPLLVDNPKVEG 109  
 QY 111 ADRLCNLFGAEKYLGGLDYAVVDFGTSTNFDVVGRRRFLGILATGAQVSADALFARA 170  
 DB 110 ADRIVNCCLAA--YDRFRKAAIVDFGSSICVDVVSAGKEFLGAGTAPGVQVSSDAAARS 167  
 QY 171 AKLPRTITQAPETAIGKNTVHALQSLGVFGAEVMDGLLRIRRAELPG-----EAVAVAT 225  
 DB 168 AALRRVELARPSVVGKNTVECMQAGVFGAGLVGLVGRIDREYVSGFSVDHDAIVAT 227  
 QY 226 GGFSTVQICQEDYDDETTLRGLVLMASRSEVR 262  
 DB 228 GHTAPLLPELHTVDHQHLLTQGLRLVFERNLEVQ 264  
 RESULT 11  
 AAU91173  
 ID AAU91173 standard; Protein; 260 AA.  
 XX  
 AC AAU91173;  
 XX  
 XX 05-JUN-2002 (first entry)  
 DT Pantothenate kinase (Coax) #11.  
 DE



XX KW Pantothenate kinase; Coax; antibiotic; antimicrobial;  
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.  
 XX Clostridium difficile.  
 OS WO200216601-A2.  
 XX PD 28-FEB-2002.  
 XX 24-AUG-2001; 2001WO-US26531.  
 XX 24-AUG-2000; 2000US-227860P.  
 PR 20-MAR-2001; 2001US-0813453.  
 XX (OMNI-) OMNIGENE BIOPRODUCTS INC.  
 PA Yocum RR, Patterson TA;  
 PI WPI; 2002-269358/31.  
 XX N-PSDB; ABK54196.  
 DR  
 PT Identifying potential antibiotic or antimicrobial agent, comprises  
 PT contacting composition comprising pantothenate kinase (Coax) protein  
 PT with test compound and identifying inhibitor of the Coax protein -  
 XX Claim 6; Page 105; 128pp; English.  
 PS  
 XX The invention describes assays for identifying a (potential) antibiotic  
 CC comprising contacting an assay composition comprising a pantothenate  
 CC kinase (Coax) protein with a test compound, and determining the ability  
 CC of the test compound to inhibit the activity of the Coax protein, an  
 CC essential enzyme for the production of coenzyme A. Coax protein is a  
 CC valuable target for identifying bactericidal compounds. Coax modulating  
 CC agents can be used in an infectious animal model to determine the  
 CC efficacy, toxicity, or side effects of treatment with such an agent. This  
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
 CC described in the invention.  
 XX Sequence 256 AA;  
 SQ  
 Query Match 28.5%; Score 377; DB 23; Length 256;  
 Best Local Similarity 35.7%; Pred. No. 3.3e-31;  
 Matches 94; Conservative 51; Mismatches 102; Indels 16; Gaps 9;  
 QY 6 LLAVDIGNTTTVGLADASGALHTWIRTNREMLPDDALQLHGLFTLAGAPIRA--A 63  
 DB 2 LLVFDVGNNTNVLGIYKGD-KLVNRYWIKTDREKTSDEYGLISNLFYDYNVISDIDV 60  
 DB 64 VLSSVAPPVGG---ENYALA-LKRHEMIDAFVSAENLPDVTVELDTPGSGADRLCN-LF 118  
 DB 61 IISVVPPNVMHSLNFCIKYCKKQPLIYVPGIKT----GLNKKYDNPKQVGNADRTVNVA 116  
 QY 119 GAERYLGLDYAVVVDFTSTNFDVVGRRFLGGILATGAQVSADALFARAALKPRITL 178  
 DB 117 GIEKY--GAP-SILVDFGTATTCFAISEKGYLGCTIAPGKISSEALFQSGASKUPRVEL 173  
 QY 179 QAPETAIGNVTHALQSLGVGYAEVMDGLLRRAELPGEAV-AVATGGFSRTVQIGCQ 237  
 DB 174 AKPGMTICKSTVSAQSGIYGVGLVDKIIISMKKELNCDDVKVIATGGLAKLIASETK 233  
 QY 238 EIDYDETLLRLGLVELWASKE 260  
 DB 234 SIDYVDGFLTLEGLRIIYEKNOE 256  
 RESULT 14  
 ID ABB47661  
 XX ABB47661 standard; Protein; 259 AA.  
 AC ABB47661;  
 XX  
 DT 05-FEB-2002 (first entry)

XX DE Listeria monocytogenes protein #365.  
 XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 XX OS Listeria monocytogenes.  
 XX PN WO200177335-A2.  
 XX PD 18-OCT-2001.  
 XX 11-APR-2001; 2001WO-FR01118.  
 XX 11-APR-2000; 2000FR-0004629.  
 PR (INSP ) INST PASTEUR.  
 PA Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 DR  
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides -  
 XX Claim 6; SEQ ID No 366; 192pp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms,  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 259 AA;  
 SQ  
 Query Match 27.5%; Score 363.5; DB 23; Length 259;  
 Best Local Similarity 31.7%; Pred. No. 8.8e-30;  
 Matches 80; Conservative 53; Mismatches 114; Indels 5; Gaps 4;  
 QY 6 LLAVDIGNTTTVGLADASGALHTWIRTNREMLPDDALQLHGLFTLAG-APIP-RAA 63  
 DB 2 ILVIDVGNNTCTGVGYEKQKLLKH-WRMTDRHTSDLGMTVLFNFYSANLTPSDTOGI 60  
 QY 64 VLSSVAPPVGENYALAKRHEMIDAFVSAENLPDVTVELDTPGSGADRLCNLFGEAKY 123  
 DB 61 IISVVPPNVMHMETMTCVRYFNIRPLIVGPGIKTGLNLRKVDNPRETSGDRIVNAVASEE 120  
 QY 124 LGGLDYAVVVDFTSTNFDVVGRRFLGGILATGAQVSADALFARAALKPRITLOAPET 183  
 DB 121 YG--TPVIVDFTGATTTCYIDESGVYQGAATAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
 QY 184 AIGKNTVHALQSLGVGYAEVMDGLLRRAELPGEAVATGGFSRTVQIGCQEIYD 243

Db 179 IIGSTVSSMOAGIFYGVGOCEGIIAEMKKQSNASPVVATGGGLARMITEKSSAVDILD 238  
Qy 244 ETTLRLGLVELW 255  
Db 239 PFLTLKGLLELY 250

RESULT 15

AAU91153  
ID AAU91153 standard; Protein; 258 AA.

XX AC AAU91153;

DT 05-JUN-2002 (first entry)

XX Rhodobacter capsulatus pantothenate kinase Coax.

DE Pantothenate kinase; Coax; antibiotic; antimicrobial;  
KW Pantothenate kinase modulator; coenzyme A; bactericidal compound.

OS Rhodobacter capsulatus.

XX WO200216601-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26531.

XX PR 24-AUG-2000; 2000US-227860P.

XX PR 20-MAR-2001; 2001US-0813453.

XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX PI Yocum RR, Patterson TA;

XX DR WPI; 2002-269358/31.

XX Identifying potential antibiotic or antimicrobial agent, comprises  
PT contacting composition comprising pantothenate kinase (Coax) protein  
PT with test compound and identifying inhibitor of the Coax protein -

PS Claim 10; Page 71-72; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic  
CC comprising contacting an assay composition comprising a pantothenate  
CC kinase (Coax) protein with a test compound, and determining the ability  
CC of the test compound to inhibit the activity of the Coax protein, an  
CC essential enzyme for the production of coenzyme A. Coax protein is a  
CC valuable target for identifying bactericidal compounds. Coax modulating  
CC agents can be used in an infectious animal model to determine the  
CC efficacy, toxicity, or side effects of treatment with such an agent. This  
CC is the amino acid sequence of a pantothenate kinase (Coax) protein  
CC described in the invention.

XX SQ Sequence 258 AA;

Query Match 26.2%; Score 347.5; DB 23; Length 258;  
Best Local Similarity 36.0%; Pred. No. 4.3e-28;  
Matches 90; Conservative 42; Mismatches 109; Indels 9; Gaps 6;  
Qy 6 LLAVDIGNTTVLGLADASGALHTWIRITNREMLPDDIALQLHGLFTLAGAP-IPRAAV 64  
Db 2 LLCIDGNTNTVFSVWDGTD-PAATWRIATDHRRTADEYFVWLTMLQLKGLGRISEAI 60  
Qy 65 LSSVAPPVGGENTALKRHFMDAFVSAE--NLDPVTVELTPGVSAGADRLCN-LFGAE 121  
Db 61 ISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELP-VAPRVDPGTTVGPDRLVNTVAGYD 119  
Qy 122 KYLGLDYAVVVDFTSTNFDVVGRCRFLGILATGAQVSADALFAAKLPRTILOAP 181  
Db 120 REGGDL---IVYDFGTATFDVVPAGYIGGVYAPGNLSLEALHMAAALPHVDVTKP 176

Qy 182 ETAIGKNTVHALQSLGVFGYAEWVDGLLRIRRAELPGEAVAVATGFSRTVOGICQEIY 241  
Db 177 QGVICTNTVACIQSGVTWGYIGLVEGIYVQIRMERDRPMKVIATGGLASLFDLGFDFDK 236  
Qy 242 YDETLTLRGL 251  
Db 237 VEDDLTMHGL 246

Search completed: June 24, 2003, 21:46:16  
Job time : 27.899 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.55623 seconds  
(without alignments)  
2635.685 Million cell updates/sec

Title: US-09-813-453A-8  
Perfect score: 1324  
Sequence: 1 MPAPPLAVDIGNTTTLVL.....DETTLRGLVELWASRSEVR 262

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1324	100.0	262	2 E75516	conserved hypothet
2	437.5	33.0	254	2 F83660	hypothetical prote
3	409.5	30.9	265	2 T36391	hypothetical prote
4	404	30.5	272	2 A70955	hypothetical prote
5	401	30.3	274	2 H86937	conserved hypothet
6	396.5	29.9	261	2 B87489	transcription acti
7	378.5	28.6	233	2 S66100	conserved hypothet
8	371.5	28.1	273	2 E97293	probable transcrip
9	363.5	27.5	259	2 AF1102	conserved hypothet
10	356.5	26.9	259	2 AF1464	conserved hypothet
11	260	19.6	246	2 D73220	conserved hypothet
12	248	18.7	273	2 D71326	conserved hypothet
13	159	12.0	262	2 F70165	conserved hypothet
14	152.5	11.5	276	2 A12292	hypothetical prote
15	133.5	10.1	229	2 E70465	hypothetical prote
16	128.5	9.7	242	2 A82637	conserved hypothet
17	123.5	9.3	257	2 S75559	hypothetical prote
18	104.5	7.9	400	2 T29121	hypothetical prote
19	103.5	7.8	267	2 I40327	baf protein - Bord
20	102.5	7.7	675	2 A75627	H+/K+-exchanging A
21	101.5	7.7	444	2 D87557	major facilitator
22	101	7.6	636	2 T35182	probable ABC-type
23	101	7.6	1441	2 B86807	hypothetical prote
24	100.5	7.6	592	2 H82031	probable biotin-[a
25	99.5	7.5	544	2 H72647	hypothetical prote
26	98.5	7.4	302	2 G70614	hypothetical prote
27	96.5	7.3	515	2 AF3524	hypothetical prote
28	96.5	7.3	592	2 B81009	BirA protein/Bvg a
29	96	7.3	360	2 S06287	fragmin - slime mo

30	94.5	7.1	314	2 H97764	malate dehydrogena
31	94	7.1	485	2 F87448	succinylglutamic s
32	93.5	7.1	419	2 D69202	5-enolpyruvylshiki
33	91.5	6.9	998	2 G83022	probable two-compo
34	91	6.9	439	2 F84278	3-phosphoshikimate
35	90.5	6.8	258	1 G69985	probable enoyl-CoA
36	90.5	6.8	317	2 F72246	glucokinase - Ther
37	90.5	6.8	346	2 T31263	xylene monooxygena
38	90.5	6.8	628	2 AC2261	cell division prot
39	90	6.8	519	2 A10400	glutamate-cysteine
40	89.5	6.8	332	2 A86892	glycosyltransferas
41	89.5	6.8	405	2 H81995	probable transmemb
42	89.5	6.8	499	2 E82977	glycolate oxidase
43	89.5	6.8	503	2 T37119	probable membrane-
44	89.5	6.8	530	2 AF2175	hypothetical prote
45	89.5	6.8	876	2 D70971	hypothetical prote

ALIGNMENTS

RESULT 1

E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R:White, O.; Eisen, J.A.; Heidberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
M.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WHI>  
A:Cross-references: GB:AE001905; GB:AE000513; NID:96458144; PIDN:AAF10040.1; PID:96458144  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match	100.0%	Score 1324;	DB 2;	Length 262;
Best Local Similarity	100.0%	Pred. No. 9.1e-101;		
Matches 262;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MPAPPLAVDIGNTTTLVLGLADASGALHTHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP	60	
Db	1	MPAPPLAVDIGNTTTLVLGLADASGALHTHTWIRTNREMLPDDDLALQLHGLFTLAGAPIP	60	
QY	61	RAVLSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA	120	
Db	51	RAVLSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGA	120	
QY	121	EKYLGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSDALFARAALPRITLQA	180	
Db	121	EKYLGLDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSDALFARAALPRITLQA	180	
QY	181	PETAIGNKTVHALQSLGVFGYAEVMDGLLRRAELPGEAVAVATGFSRTVQICGEID	240	
Db	181	PETAIGNKTVHALQSLGVFGYAEVMDGLLRRAELPGEAVAVATGFSRTVQICGEID	240	
QY	241	YDETTTLRGLVELWASRSEVR	262	
Db	241	YDETTTLRGLVELWASRSEVR	262	

RESULT 2

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans



A:Accession: H86937

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-274 <STO>

A:Cross-references: GB:AL450380; NID:gl3092576; PIDN:CAC29740.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML0232

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.3%; Score 401; DB 2; Length 274;  
Best Local Similarity 38.8%; Pred. No. 2.3e-25;  
Matches 104; Conservative 38; Mismatches 90; Indels 36; Gaps 7;

QY 6 LLAVIDGNTTTLGLADASGALHTHT-----WRIRTNREMLPDDALQLHGLFTLAGAPIP 60

DB 2 LLAIDVRNTHTVGL--LGSSEKHAQVQWRIETSEVTADELALIDGLDDSERLA 59

QY 61 RAAVLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVE-----LDPGSGVG 110

DB 60 GAAALSTV-PSVLEHVRMLDQW-----PSVPHVLEPGVVRTGIPLLVDNPKYVG 109

QY 111 ADRLCNLFGEAKYLGGLDYAVVVDFGTSTNFDVGRGRFLGILATGAQVADALFARA 170

DB 110 ADRIVNCALAAHKFG--QAAIVYDFGSSICVDVSAKGEFLGALAPGVQVSSDRAARS 167

QY 171 AKLPRITLQAPETAIGKNTVHALQSLGVFGYAEVMDGLLRIR-----AELPGEAVAV 223

DB 168 AALRRVELARPSVVGNKNTVECMQAGVFGFAGLVGLVGRMRQDVVEEFGDLGNRAVV 227

QY 224 ATGGSRTVVGICQIDYDETLTLRGL 251

DB 228 ATGHTAPLPLLEHTVVDYDRLTLHGL 255

#### RESULT 6

B87489

Transcription activator, probable Baf family [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: B87489

R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Lau, B.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87489

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <STO>

A:Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1935

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 29.9%; Score 396.5; DB 2; Length 261;  
Best Local Similarity 36.1%; Pred. No. 5e-25;  
Matches 92; Conservative 45; Mismatches 109; Indels 9; Gaps 5;

QY 6 LLAVIDGNTTTLGLADASGALHTHTWRIRTNREMLPDDALQLHGLFTLAGAPIP--AA 63

DB 3 LLAIEQGNNTMFAIHGASVQAQ--WRSATSTRTADYVYVWLSQLLSQGLGFRADAV 61

QY 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGVADRLCNLFGEAK- 122

DB 62 IISVVVQSFIFNLRLNLSRYFNVEPLVIGENAKGLIDVRIEKPSEAGADRLVNAIGAAMV 121

QY 123 YLGGILDYAVVVDFGTSTNFDVGRGRFLGILATGAQVADALFARAALPRITLQAP- 181

DB 122 YGGLP---VVIDSGTATTFDVAADGAFEGGIIITAPGINLSQALHAAKLPRIATQFPA 178

QY 182 -ETAIGKNTVHALQSLGVFGYAEVMDGLLRIRRAELPGEAVAVATGSGFSTVVGICQID 240

DB 181 -ETAIGKNTVHALQSLGVFGYAEVMDGLLRIRRAELPGEAVAVATGSGFSTVVGICQID 240

DB 179 GNRIVGTDTVSAMQSGVFWGYSISLIEGLVARIKARGEPMVTIATCGVASLFEAGTDSID 238

QY 241 YYDETTLTLRGLVELW 255

DB 239 HFDSDLTIRGLLEIY 253

#### RESULT 7

S66100

Conserved hypothetical protein yacB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000

C:Accession: S66100; E69740

R.Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994

A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis

A:Reference number: S65967; MUID:96051385; PMID:7584024

A:Accession: S66100

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <OGA>

A:Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g41

R.Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; B.

C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; G.

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardi,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; M.

y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Porc

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Sca

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.;

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchi

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshi

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. Bacillus subtil

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: E69740

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <KUN>

A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el18

A:Experimental source: strain 168

C:Genetics:

A:Gene: yacB

A:Start codon: TTG

C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.6%; Score 378.5; DB 2; Length 233;  
Best Local Similarity 38.3%; Pred. No. 1.3e-23;  
Matches 82; Conservative 44; Mismatches 75; Indels 13; Gaps 5;

QY 6 LLAVIDGNTTTLGLADASGALHTHTWRIRTNREMLPDDALQLHGLFTLAGAPIP--AA 63

DB 2 LLYVDGNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMLRSLDFSLGIMFEQIDGI 60

QY 64 VLSSVAPPVGENYALAKR---HEMIDAFVSAENLPDVTVELDTPGSGVADRLCNLF 119

DB 61 IISVVVPPI---MFALERCTKYFHIEPQIVGPMGTGLNKNKYDNKFEVGAADRIYNVA 116

QY 120 AEKYLGLDYAVVVDFGTSTNFDVGRGRFLGILATGAQVADALFARAALPRITLQ 179

DB 117 AIHLYG--NPLIVVDFGTATYCYIDENKQYMGGAIPGITISTEALYSRAKLPRIET 174

QY 180 APETAIGKNTVHALQSLGVFGYAEVMDGLLRIR 213

DB 175 RPDNIIGKNTVSAMQSGILFGYGVQVEGIVKRMK 208

#### RESULT 8

E97293

probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridium  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: E97293  
R.; Nolling, J.; Brennon, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97293  
A:Status: preliminary  
A:Residues: 1-273 <KUR>  
A:Molecule type: DNA  
A:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:gl5026270; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC3200  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.1%; Score 371.5; DB 2; Length 273;  
Best Local Similarity 34.5%; Pred. No. 5.8e-23;  
Matches 89; Conservative 51; Mismatches 97; Indels 21; Gaps 6;

QY 6 LLAVDIGNTTVGLADASGALHTWRIRTNREMLPDDLALQLHGLF-----TLGAP 58  
DB 13 ILVLDVGNNTVILGIYNDP-KLTAEWRLSTDLVLSADEYCIQVWNLFFQDKLPTLV--- 68  
QY 59 IPRAAVLSSVPPVGENYALAKRHMIDAFVSAENLPDVTVELTPGSGVADRLCNLF 118  
DB 69 --EGVISSVWPNIMYSLEHMRKFKYKINPLVPGIKTGINKIKNPKEVGDRIYNAV 126  
QY 119 GA-EKYLGLDVAVVDFGTSNFDVVGRRFLGGLATGAQVSADALFARAALPRIT 177  
DB 127 AHEIYKRL-----IIIDFGTATTCFAVRENGDYLGAICPGIKVSEALFEKAALPRVE 183  
QY 178 LQAPETAIGNKTVHALQSLGVFAEMVDGLLRIRAEPLPGEA---VAVATGGFGRVTQ 233  
DB 184 LIKPAVAICKNTISSIQSGIVGYIGQVRYIVRMKEELQEGEKEPLVAVATGGLAKLIS 243  
QY 234 GICQEDIDYDEILTLRGL 251  
DB 244 EEAKNVVDVINPFLTEGL 261

RESULT 9  
AF1102  
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
Accession: AF1102  
R.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative Genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1102  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00748.1; PID:gl6409586; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0221  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 27.5%; Score 363.5; DB 2; Length 259;  
Best Local Similarity 31.7%; Pred. No. 2.4e-22;  
Matches 80; Conservative 53; Mismatches 114; Indels 5; Gaps 4;

QY 6 LLAVDIGNTTVGLADASGALHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63

Db 2 ILVIDGNTNCTGVYKEQKLLKH-WRMTTDRHRTSDELGMTVLNFFSYANLTPSDIOGI 60  
QY 64 VLSSVAPPVGENYALAKRHMIDAFVSAENLPDVTVELTPGSGVADRLCNLFGAERY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGILKLVNDNPREIGSDRIVNAVAASEE 120  
QY 124 LGGLDYAVVDFGTSNFDVVGRRFLGGLATGAQVSADALFARAALPRITLOAPET 183  
Db 121 YG--TPVIVVDFGTATTCFYIDESGVYGGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
QY 184 AIGKNTVHALQSLGVFAEMVDGLLRIRAEPLPGEA VAVATGGFGRVTQICQIDYD 243  
Db 179 IIGKSTVSSVQAGIFGFGVQCCEGIIAEMKKQSNASPVVATGATGLARMIKTEKSSAVDILD 238  
QY 244 ETTLRGLVELW 255  
Db 239 PFLTLKGLLELY 250

RESULT 10  
AF1464  
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1464  
R.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi  
Science 294, 849-852, 2001  
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weh  
D.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:gl6412682; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.9%; Score 356.5; DB 2; Length 259;  
Best Local Similarity 31.7%; Pred. No. 9.1e-22;  
Matches 80; Conservative 52; Mismatches 115; Indels 5; Gaps 4;

QY 6 LLAVDIGNTTVGLADASGALHTWRIRTNREMLPDDLALQLHGLFTLAG-APIP-RAA 63  
Db 2 ILVIDGNTNCTGVYKEQKLLRH-WRMTTDRHRTSDELGMTVLNFFSYANLTPSDIOGI 60  
QY 64 VLSSVAPPVGENYALAKRHMIDAFVSAENLPDVTVELTPGSGVADRLCNLFGAERY 123  
Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVPGIKTGILKLVNDNPREIGSDRIVNAVAASEE 120  
QY 124 LGGLDYAVVDFGTSNFDVVGRRFLGGLATGAQVSADALFARAALPRITLOAPET 183  
Db 121 YG--TPVIVVDFGTATTCFYIDEGVYGGGAIAPGIMISTEALYNRAAKLPRVDIAESSQ 178  
QY 184 AIGKNTVHALQSLGVFAEMVDGLLRIRAEPLPGEA VAVATGGFGRVTQICQIDYD 243  
Db 179 IIGKSTVSSVQAGIFGFGVQCCEGIIAEMKKQSNASPVVATGATGLARMIKTEKSSAVDILD 238  
QY 244 ETTLRGLVELW 255  
Db 239 PFLTLKGLLELY 250

RESULT 11  
D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima

RESULT 13  
F70165  
conserved hypothetical protein BB0527 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: F70165  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra,  
son, D.; Peterson, J.; Kertavagae, A.R.; Quackenbush, J.; Salzberg, S.; Hanson,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Ha  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: F70165  
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <RLE>  
A:Cross-references: GB:AE001154; GB:AE000783; NID:Q2689431; PIDN:AAC66892.1; P  
A:Experimental source: strain B31

Query Match	12.0%	Score 159;	DB 2;	Length 262;
Best Local Similarity	23.2%	Pred. No. 1.2e-05;		

QY	2	PAPFLLAVDIGNTTTLVLGADASGALTHWTIRTNREMLPDDLLALQLHGLFT	LACAPIR	61
		- : : : : -		
		- : : : : -		
Dh	4	PII:SEVITDIGNTSIAPALF-KNOVNIPIKMKTNLMRYDVEFFEEENFDN----	VN	58

QY	62	AAVLSSVAPPGENYALAKRHHMIDAFVSAENLPDVTE--LDTPGSVGADRLCNLF	119
		: : : : : : : : : : : : : : : : :	
Db	59	KVRISSVPIINETFKNWFISFFRIKPLFICGFDLNYDITFNPYKSKDFLLGSDVFANLVA	118

Qy		120	A-EKYLGLDYAVVVDSEGTSTNFVDVGRRFLGGILATGAQVSADALFAARAAKLPRIITL	178
Dd		119	AIENY-SFENVLVVDLGTACTIFAVSRQDGLGGIINSPLINSLDNAYLKPKFI	176
Qy		179	QAPETAIGKNTVHALQSGLVFGEAEMVDGLRRLRAELPGEAVAVATGFSRTVOICOE	238
Dd		177	STPNLLERTSGSVNSGLFYQKYKLYEGVYDIKQMYKKFNLIITGNADLLISLI-E	235
Qy		239	IDY-YDETLTRLRGL	251
Dd		236	IEFFNIHLTVEGV	249

RESULT 14  
AI2292  
hydrothermal protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)

R.; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; DNA Res. 8, 205-213, 2001

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AI2292  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-276 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA075595.1; PID:g17133030; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: a1r3896

Query Match 11.5%; Score 152.5; DB 2; Length 276;  
Best Local Similarity 24.4%; Pred. No. 4.5e-05;  
Matches 54; Conservative 41; Mismatches 73; Indels 53; Gaps 10;  
QY 7 LAVDIGNTTIVLGLADASGALHTWRTNREMLPDDLALQLH----- 50  
DB 15 LALEIGNSLRWALF-MGESLETFW---DTEYLPESVIQOLNGETKLEVGSEKEIFF 69  
QY 51 -LFTLAGAPTPRAAVLSSVAPP---VGENYALAKRHFMDAFVSAENLPDVTVELDTP 106  
DB 70 TTFPLPAPCPPLFLFTASVVPQOTVLWNY-----LNVRVITLQDIPLNIIY--- 116  
QY 107 GSVGADRLCNLFCAERYLGLDY---AVVDFGTSTNFDVVGRRFLGGLATGAQVSA 163  
DB 117 PTLGIDRALALNGA-----GMSWGFPPVLIDAGTALTFTAADGCKNLVGGAILPGVGLQF 171  
QY 164 DALFARAALPRITLQA-----PETAIGNVTVALOSGLVF 199  
DB 172 ASLIGQOTGOLPOVEMEAIRKSLPPRFAL--NTTEAIOQGVII 210

RESULT 15

E70465  
hypothetical protein aq\_1924 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: E70465  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70465  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-229 <AQF>  
A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE000059  
A:Experimental source: strain VF5  
C:Genetics:  
Gene: aq\_1924

Query Match 10.1%; Score 133.5; DB 2; Length 229;  
Best Local Similarity 23.0%; Pred. No. 0.0013;  
Matches 58; Conservative 44; Mismatches 113; Indels 37; Gaps 9;  
QY 7 LAVDIGNTTIVLGLADASGALHTWRTNREMLPDDLALQLHGLFTLAGAPTPRAAVLS 66  
DB 4 LTVDVGNSSVDIALWE--GKKVKDFLKLSHSEFLKEE-----PPK---LK 43  
QY 67 SVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFCAERYLGG 126  
DB 44 ALGISVKSFSSEKVKIPKIF-LKKENFP-IQVDYKTPETLGTDRVALAYSAKKFFYG- 100  
QY 127 LDYAVVDFGTSTNFDVVGRRFLGGLATGAQVSAADALFARAALPRITLQAPETAIG 186  
DB 101 -KNVVISAGTALVIDLVLEG-KFKGGFTLGLGKKLKLSDLAEGIPPEFFEEVIFLG 158  
QY 187 KNTVALOSGLVFGYAEAMVDGLLRIRIARLPGELVAVATG-----FSRTVQIGCOIDYY 242  
DB 159 RSTRECVLGGAYRESTEFIKSTLKLWRKVKRKKVITGGEKVFYSK-----FGIY 210  
QY 243 DETLRLGLVEL 254

Db 211 DPLLVRHGMENL 222

Search completed: June 24, 2003, 22:03:05  
Job time : 11.5562 secs

US-09-813-453A-5  
; Sequence 5, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
; -09-813-453A-5

Query Match 29.4%; Score 382.5; DB 9; Length 272;  
Best Local Similarity 34.0%; Pred. No. 2.1e-28;  
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY 1 MLLAIEOGNTNTMFAIHG-----ASWVAQWRSATSTRTADYVYVWLSQLLSMOGLGFRA 56  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 57 IDAVIISVVVPSQIFNLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 57 IDAVIISVVVPSQIFNLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 116 GAAMVYGPPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIOR 175  
Db 58 LTGTAALSTVPSVLHEVRIMLDQWPSVPHVLEPGVRTGIPLLVDNPKVEGADRVNCL 117

QY 176 PAGNRIVGTDTVSAMQSGVFWGYSISLIEGLVARIKAEKRG-----EPMVTIATGGVASLFE 230  
Db 178 P--RSVVGKNTVECMQAGAVFGAGLVGLVGRIREDSVGFSDHVDHVAIVATGHTAPLLL 235

QY 231 GATDSIDHFDSDLTIRGLLEIYRN 255  
Db 236 PELHTVDHYDQHLTQGLRLVFN 260

Search completed: June 24, 2003, 22:29:03  
Job time : 15.4494 secs

US-09-813-453A-5  
; Sequence 5, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
; -09-813-453A-5

Query Match 29.4%; Score 382.5; DB 9; Length 272;  
Best Local Similarity 34.0%; Pred. No. 2.1e-28;  
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY 1 MLLAIEOGNTNTMFAIHG-----ASWVAQWRSATSTRTADYVYVWLSQLLSMOGLGFRA 56  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 57 IDAVIISVVVPSQIFNLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 57 IDAVIISVVVPSQIFNLNLSRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115  
Db 1 MLLAIDVRNTHTVVGLLSGKHEAKVYVQWRIETSEVTDELALATIDGLI---GEDSER 57

QY 116 GAAMVYGPPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIOR 175  
Db 58 LTGTAALSTVPSVLHEVRIMLDQWPSVPHVLEPGVRTGIPLLVDNPKVEGADRVNCL 117

QY 176 PAGNRIVGTDTVSAMQSGVFWGYSISLIEGLVARIKAEKRG-----EPMVTIATGGVASLFE 230  
Db 178 P--RSVVGKNTVECMQAGAVFGAGLVGLVGRIREDSVGFSDHVDHVAIVATGHTAPLLL 235

QY 231 GATDSIDHFDSDLTIRGLLEIYRN 255  
Db 236 PELHTVDHYDQHLTQGLRLVFN 260

Search completed: June 24, 2003, 22:29:03  
Job time : 15.4494 secs

## .RESULT 15

3-09-712-363-276  
Sequence 276, Application US/09712363  
Patent No. US20020164588A1  
GENERAL INFORMATION:  
APPLICANT: Eisenberg, David  
APPLICANT: Rotstein, Sergio H.  
APPLICANT: Marcotte, Edward M.  
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
FILE REFERENCE: 07419-032001  
CURRENT APPLICATION NUMBER: US/09/712,363  
CURRENT FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: PCT/US00/02246  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/179,531  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/117,844  
PRIOR FILING DATE: 1999-01-29  
PRIOR APPLICATION NUMBER: 60/118,206  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: 60/126,593  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 60/134,093  
PRIOR FILING DATE: 1999-05-14

```
QY 181 IVGDTVSAMQSGVFWGYISLIEGLVARIK 210
Db 179 IIGKNTVSAMQSGILFYGVQVEGIYKRMK 208

RESULT 11
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

Query Match 36.8%; Score 478; DB 9; Length 219;
Best Local Similarity 44.9%; Pred. No. 1.4e-37;
Matches 96; Conservative 41; Mismatches 75; Indels 2; Gaps 1;

QY 2 LLAIEGNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMOGLGFRADAVI 61
Db 5 LVAVDIGNTSNIGIPEGEKLLANHLGWSVAQRMADAYASLLGLLQHAGIHPLELNVI 64
QY 62 ISSVVPQSIFNLNLSRRYFNVPLVIGENAKIGIDVRIEKPSEAGADRLVNAIGAAMY 121
Db 65 MCSVVPPLTTTEEVFKSTFKAPLVVVGAGIKSGVKVRMDNPREVGADRIVNAARVLY 124
QY 122 PGLVVDGSTATTFFDIVAADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIORPAGNRI 181
Db 125 PGACIIVDMGTATFTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPA--KV 182
QY 182 VGTDTVSAMQSGVFWGYISLIEGLVARIKAEERGE 215
Db 183 IGSNTVSAMQSGIFYGYIGLVEELVRRITQELGQ 216

RESULT 12
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum

US-09-813-453A-3
Query Match 35.5%; Score 460.5; DB 9; Length 250;
Best Local Similarity 39.5%; Pred. No. 7.7e-36;
Matches 92; Conservative 53; Mismatches 81; Indels 7; Gaps 2;

QY 1 LLAIEGNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMOGLGFRADAV 60
Db 18 VILVDVGNNTIVLGIYNDTKLTAEWRLSTDVLSRSDAYEGIQVMNLFQODKLDPLVEGV 77
QY 61 IISVVPQSIFNLNLSRRYFNVPLVIGENAKIGIDVRIEKPSEAGADRLVNAIGAAMY 120
Db 78 IISVVPNIMYSLEHMIRKFKINPLVVGPGIKTINIKYDNPKVEGADRIVNAVAHEI 137
QY 121 YPGPLVVDGSTATTFFDIVAADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIORPAGN 180
Db 138 YKRSLLIIDFGTATTFCAVRENGDYLGAICPGIKVSSEALFEKAALKPRVELIKPA--Y 195
QY 181 IVGDTVSAMQSGVFWGYISLIEGLVARIK-----AERGEPMVTIATGGVASL 228
Db 196 AICKNTISSQSGIVVYRLQVKYLFKELKENLPGRETRTSLVLATGGLAKL 248

RESULT 13
US-09-813-453A-8
; Sequence 8, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

Query Match 30.5%; Score 396.5; DB 9; Length 262;
Best Local Similarity 36.1%; Pred. No. 9.5e-30;
Matches 92; Conservative 45; Mismatches 109; Indels 9; Gaps 5;

QY 2 LLAIEGNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMOGLGFRADAV 60
Db 6 LLAVDIGNTTVLGLADASGALTHTWRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63
QY 61 IISVVPQSIFNLNLSRRYFNVPLVIGENAKIGIDVRIEKPSEAGADRLVNAIGAAMY 120
Db 64 VLSVAPPVGENYALAKRHFMDAFVAVSAENLPDVTVELDTPGSGADRLCNLFGEAK- 122
QY 121 YPGPL--VVIDSGTATTFDIVAADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIORPA 177
Db 123 YLGLDYAVVVDGTSTNFDVVGRRFLGGLATGAQVSADALFARAKLPRIYLOAP- 181
QY 178 GNRIVGDTVSAMQSGVFWGYISLIEGLVARIKAEERGEPMVTIATGGVASLFEAGATSD 237
Db 182 -ETAIGKNTVHALQSGLVFGYAEVMDGLLRIRRAELPGEAVAVATGGSRTVQGCQCID 240
QY 238 HFDSDLTIRGLLEYI 252
Db 241 YYDETTLTIRGLVEIW 255

RESULT 14
```





Db 179 IIGKNTVSAMQSGILFGYGVGVEGIVKRMKQAKQDLKVIATGGGLAPLIANESDCIDIWD 238  
QY 241 SDLTIRGLLEIYRNTI 257  
Db 239 PFLTKGLGLEIYERNRV 255

## RESULT 5

US-09-813-453A-49  
; Sequence 49, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-813-453A-49

Query Match 44.3%; Score 575; DB 9; Length 258;  
Best Local Similarity 44.1%; Pred. No. 1.1e-46;  
Matches 113; Conservative 52; Mismatches 89; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDSASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKSTDEYDGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPOSIFNLRLNLSRRYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPPIMPFALEMRCKYFKIPLVVGPGIKTGLNIRKYPENPREVGADRLVNAVAGIHL 120  
QY 121 YPGPLVWIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIORPAGNR 180  
Db 121 YGSPLIIVDFGTATTCYINEKHQYMGAIAPGIMISTEALFARAALPRIEIRP--DD 178  
QY 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASLFEGATSDIDHFD 240  
Db 179 IIGKNTVSAMQAGILYGVGVEGIVSRMKAISKIPKVIATGGGLAPLIASESDIIDVD 238  
QY 241 SDLTIRGLLEIYRNT 256  
Db 239 PFLTKGLKLLYKENT 254

## RESULT 6

US-09-813-453A-45  
; Sequence 45, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21

; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 262  
; TYPE: PRT  
; ORGANISM: Bacillus anthracis  
US-09-813-453A-45

Query Match 43.6%; Score 566; DB 9; Length 262;  
Best Local Similarity 42.0%; Pred. No. 8.3e-46;  
Matches 107; Conservative 55; Mismatches 91; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDSASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MIFVLDVGNNTVGLVYDGDDELKHHWRIETSRKSTDEYDGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPOSIFNLRLNLSRRYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPPIMPFALEMRCKYFKIPLVVGPGIKTGLNIRKYPENPREVGADRLVNAVAGIHL 120  
QY 121 YPGPLVWIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIORPAGNR 180  
Db 121 YGSPLIIVDFGTATTCYINEKHQYMGAIAPGIMISTEALFARAALPRIEIRP--DD 178  
QY 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVASLFEGATSDIDHFD 240  
Db 179 VVGKNTVSAMQSGILYGVGVEGIVKRMKEAKQEPKVIATGGGLAKLISESNVIDVD 238  
QY 241 SDLTIRGLLEIYRNT 255  
Db 239 PFLTKGLYMLYERN 253

## RESULT 7

US-09-813-453A-7  
; Sequence 7, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OGZ-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 255  
; TYPE: PRT  
; ORGANISM: Geobacter sulfurreducens  
US-09-813-453A-7

Query Match 43.2%; Score 561; DB 9; Length 255;  
Best Local Similarity 44.7%; Pred. No. 2.4e-45;  
Matches 114; Conservative 50; Mismatches 89; Indels 2; Gaps 1;  
QY 1 MLLAIEQGNNTMTFAIHDSASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60  
Db 1 MLLVIVDGNNTVGLVYDGDDELKHHWRIETSRKSTDEYDGMKALLNHVGLQFSDIRGI 60  
QY 61 IISVVPOSIFNLRLNLSRRYFNVEPLVIGENAKLGDIVRIEKPSEAGADRLVNAIGAAMV 120  
Db 61 IISVVPPIMPFALEMRCKYFKIPLVVGPGIKTGLNIRKYPENPREVGADRLVNAVAGIHL 120  
QY 121 YPGPLVWIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIORPAGNR 180  
Db 121 YRTSLIIVDFGTATTFDYNRKEGCGGAIAPLVISTEALFORASKLPVRDIIRPSA-- 178

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match      46.0%; Score 598; DB 9; Length 254;
Best Local Similarity 43.9%; Pred. No. 7.4e-49;
Matches 11; Conservative 61; Mismatches 80; Indels 2; Gaps 1;

Qy 1 MLLAIEGCGNTNTMFAlHDGASWQAQRSATSESTRADEYVVMLSQLLSMOGLGFRAIDAV 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MILVIDGNTNTVLGVGYODETLVHHWRLATSRQKTEDEXAMTVRSFLDHAGLQFQDlGI 60

Qy 61 IISVVPPQSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSGAGADRLVNIAIGAAV 120
   ||||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 61 VISSVPPMFMFSLEQMCKKYFHVTPIIIGPICKTGLNIKYDNPNKEVGADRIVNVAAIL 120

Qy 121 YGPLVLVDGSTATTFFDIVAADGNAFEGGITAPGINLSMQALHEAAAKLPRIAIOQPAGNR 180
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 121 YGPFAIYVDFGTATTYCLINEKKOYAGGVITAPGIMISTEALYHRSKLPRIETIAKP -- KQ 178

Qy 181 IVGTDTSVMSQSGVFWEYGYSILIEGLVARIKAECPMTPTATGVSASFEGATSIDHF 240
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 179 VVGNTTDSMQSGIFYGVGSQDVGVVKRKAQAESEPKVIATGGLAKLIGTSETIDVID 238

Qy 241 SDLTI RGLLEIYYRN 255
   | | : | : | | | : |
Db 239 SFELTKGLQLYKKN 253
```

RESULT 4  
US-09-813-453A-2  
; Sequence 2, Application US/09813453A  
; Patent No. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; APPLICANT: Patterson, Thomas A.  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-813-453A-2

	Query Match	44.4%	Score 577;	DB 9;	Length 258;
	Best Local Similarity	44.4%;	Pred. No. 7.4e-47;		
Matches 114; Conservative		50;	MisMatches 91;	Indels 2;	Gaps 1;
QY	1 MLLAEGONTMTFAIHDSGASWAQRWTSATESTRDAEYYVVLVSOLLSSMOGLGFRAIDAV	60			
:	: : : : : :	:	:	:	:
Dd	1 LLLVIDGVNTTVLGGVDHDKLEYHWRIETSRHKTEDEFGMILKSLFPDHSGLMFPEQIDCI	60			
:	: : : : : :	:	:	:	:
QY	61 IISVVPSQSIFNLRLNRRYPVEPLVIGENAKIGDVIRIEKPSEAGADRLVLNAIGAAMV	120			
:	: : : : : :	:	:	:	:
Dd	61 IISVVPPIMPALERMCCTKYFHIEPQIVPGMKTGLNIKDNPKEVGADRIVNVAAIHL	120			
:	: : : : ~~~~~~	:	:	:	:
QY	121 YGPLVIYSDSTATTFDIVAADGAPEGGIAPGINLSMAOLHEAAAKLPRIAIORPACNR	180			
:	: : : : : :	:	:	:	:
Dd	121 YGNLVIVDFGTATVCYIDNSKNQMCGAIAPGITISTEALYSRAAKLPRIETRP--DN	178			
:	: : : : : :	:	:	:	:
QY	181 IVGTDITVSAMSGSFWFYGISLIQGLVARIKAEERGEPMPTVIATGGVASLFEGATSIDIHFED	240			

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

'OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 14.4494 Seconds  
(without alignments)  
1947.059 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNNTMFAIHGDA.....SDLTIRGLLEIYRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1299	100.0	260	9 US-09-813-453A-51	Sequence 51, Appl
2	658	50.7	258	9 US-09-813-453A-6	Sequence 6, Appl
3	598	46.0	254	9 US-09-813-453A-47	Sequence 47, Appl
4	577	44.4	258	9 US-09-813-453A-2	Sequence 2, Appl
5	575	44.3	258	9 US-09-813-453A-49	Sequence 49, Appl
6	566	43.6	262	9 US-09-813-453A-45	Sequence 45, Appl
7	561	43.2	255	9 US-09-813-453A-7	Sequence 7, Appl
8	523.5	40.3	256	9 US-09-813-453A-55	Sequence 55, Appl
9	517	39.8	265	9 US-09-813-453A-4	Sequence 4, Appl
10	492	37.9	233	9 US-09-813-453A-17	Sequence 17, Appl
11	478	36.8	219	9 US-09-813-453A-57	Sequence 57, Appl
12	460.5	35.5	250	9 US-09-813-453A-3	Sequence 3, Appl
13	396.5	30.5	262	9 US-09-813-453A-8	Sequence 8, Appl
14	382.5	29.4	272	9 US-09-813-453A-5	Sequence 5, Appl
15	378.5	29.1	272	9 US-09-712-363-276	Sequence 276, App
16	359.5	27.7	246	9 US-09-813-453A-9	Sequence 9, Appl
17	337.5	26.0	273	9 US-09-813-453A-10	Sequence 10, Appl
18	307.5	23.7	212	9 US-09-813-453A-59	Sequence 59, Appl
19	261	20.1	257	9 US-09-813-453A-53	Sequence 53, Appl

20	227	17.5	241	9 US-09-813-453A-63	Sequence 63, Appl
21	198	15.2	244	9 US-09-813-453A-41	Sequence 41, Appl
22	189.5	14.6	262	9 US-09-813-453A-11	Sequence 11, Appl
23	185	14.2	460	9 US-09-813-453A-39	Sequence 39, Appl
24	175	13.5	592	9 US-09-813-453A-22	Sequence 22, Appl
25	172	13.2	592	9 US-09-813-453A-43	Sequence 43, Appl
26	160	12.3	248	9 US-09-813-453A-20	Sequence 20, Appl
27	153.5	11.8	249	9 US-09-813-453A-61	Sequence 61, Appl
28	147	11.3	257	9 US-09-813-453A-13	Sequence 13, Appl
29	141	10.9	249	9 US-09-813-453A-70	Sequence 70, Appl
30	140.5	10.8	242	9 US-09-813-453A-65	Sequence 65, Appl
31	134.5	10.4	267	9 US-09-813-453A-15	Sequence 15, Appl
32	125.5	9.7	229	9 US-09-813-453A-12	Sequence 12, Appl
33	111	8.5	209	9 US-09-813-453A-21	Sequence 21, Appl
34	98	7.3	1536	9 US-10-092-880-2	Sequence 2, Appl
35	94.5	7.3	448	9 US-09-738-626-3991	Sequence 3991, Ap
36	93	7.2	491	9 US-09-738-626-4901	Sequence 4901, Ap
37	92	7.1	511	9 US-09-738-626-4770	Sequence 4770, Ap
38	89	6.9	206	9 US-09-997-816-4	Sequence 4, Appl
39	88	6.8	255	10 US-09-847-637B-9	Sequence 9, Appl
40	87.5	6.7	343	9 US-10-168-066-4	Sequence 11905, A
41	87.5	6.7	923	10 US-09-815-242-11905	Sequence 10582, A
42	87	6.7	242	10 US-09-815-242-10582	Sequence 13713, A
43	87	6.7	638	10 US-09-815-242-13713	Sequence 6322, Ap
44	85	6.5	398	9 US-09-738-626-6322	Sequence 2, Appl
45	85	6.5	402	10 US-09-838-564A-2	

ALIGNMENTS

RESULT 1  
US-09-813-453A-51  
; Sequence 51, Application US/09813453A  
; Patent NO. US20020168681A1  
; GENERAL INFORMATION:  
; APPLICANT: Yocum, R. Rogers  
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: ANTIBIOTICS  
; FILE REFERENCE: OG2-001  
; CURRENT APPLICATION NUMBER: US/09/813,453A  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: US 60/227,860  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/667,569  
; PRIOR FILING DATE: 2000-09-21  
; NUMBER OF SEQ ID NOS: 77  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 260  
; TYPE: PRT  
; ORGANISM: Caulobacter crescentus  
US-09-813-453A-51

Query Match	100.0%	Score	1299	DB	9	Length	260
Best Local Similarity	100.0%	Pred. No.	2.9e-115	Mismatches	0	Indels	0
Matches	260	Conservative	0				
Qy	1	MLLAIEQGNNTMFAIHGDA	SVAGQWRSATSTRTAD	EVVWLSQLLSMQGLG	FRDAIV	60	
Db	1	MLLAIEQGNNTMFAIHGDA	SVAGQWRSATSTRTAD	EVVWLSQLLSMQGLG	FRDAIV	60	
Qy	61	IISSVVPQISFNRLNLSRRYFNVEPLVTIGENAKGLIDIVRIEKPSEAGADRLNAIGAAV				120	
Db	61	IISSVVPQISFNRLNLSRRYFNVEPLVTIGENAKGLIDIVRIEKPSEAGADRLNAIGAAV				120	
Qy	121	YPGPLVVIDSTATTDFDI	VAADGAE	GGIAPGINLSMQALHEAAKLPRIALQRPAGNR		180	
Db	121	YPGPLVVIDSTATTDFDI	VAADGAE	GGIAPGINLSMQALHEAAKLPRIALQRPAGNR		180	
Qy	181	IVGDTVSAMQSGVFWGYS	ISLIEGLVARIKAR	GPMTVIATGGV	ASLFE	240	
Db	181	IVGDTVSAMQSGVFWGYS	ISLIEGLVARIKAR	GPMTVIATGGV	ASLFE	240	

Result No.	Query		Score	Match	Length	DB	ID	Description
1	98	7.5	1095	4	US-09-206-942-69	Sequence 69, Appl		
2	98	7.5	1536	1	US-08-038-682-2	Sequence 2, Appl		
3	98	7.5	1536	1	US-08-302-832-2	Sequence 2, Appl		
4	98	7.5	1536	2	US-08-530-198-2	Sequence 2, Appl		
5	98	7.5	1536	2	US-08-469-880-2	Sequence 2, Appl		
6	98	7.5	1536	2	US-08-728-470-2	Sequence 2, Appl		
7	98	7.5	1536	2	US-08-617-697-2	Sequence 2, Appl		
8	98	7.5	1536	4	US-08-719-641-2	Sequence 2, Appl		
9	98	7.5	1536	4	US-09-206-942-67	Sequence 67, Appl		
10	90	6.9	1220	4	US-09-206-942-28	Sequence 28, Appl		
11	90	6.9	1226	4	US-09-206-942-26	Sequence 26, Appl		
12	89.5	6.9	1891	2	US-08-804-227C-12	Sequence 12, Appl		
13	89.5	6.9	1891	2	US-08-804-198-6	Sequence 6, Appl		
14	85	6.5	1489	6	5183745-2	Patent No. 5183745		
15	85	6.5	1706	4	US-08-669-785-2	Sequence 2, Appl		
16	85	6.5	1794	6	5183745-6	Patent No. 5183745		
17	84.5	6.5	1705	4	US-08-669-785-4	Sequence 4, Appl		
18	83.5	6.4	659	4	US-08-894-818B-1	Sequence 1, Appl		
19	83.5	6.4	659	4	US-09-445-472-12	Sequence 12, Appl		
20	83	6.4	1180	4	US-09-206-942-65	Sequence 65, Appl		
21	83	6.4	1188	4	US-09-206-942-63	Sequence 63, Appl		
22	82.5	6.4	3170	4	US-09-036-987A-4	Sequence 4, Appl		
23	82.5	6.4	3170	4	US-09-370-700-4	Sequence 4, Appl		
24	82	6.3	1095	4	US-09-206-942-45	Sequence 45, Appl		
25	82	6.3	1101	4	US-09-206-942-43	Sequence 43, Appl		
26	81.5	6.3	4150	4	US-09-428-517-2	Sequence 2, Appl		
27	81	6.2	3729	2	US-08-804-227C-4	Sequence 4, Appl		

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/038,682  
; FILING DATE: 16-MAR-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-293  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-038-682-2

Query Match 7.5%; Score 98; DB 1; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;  
  
QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINSQALHQAALKPRI-----171  
DB 1147 ITTKGTNNATTGNVEITAGTGSILGSISSSGSVTLTAT-EGALAVSNISGNTVTVA 1205  
  
QY 172 ---AIQRPAGNRIVGDTV-SAMQSGVFWGVIYIS-----LIEGLV-----ARIKAERGE 216  
DB 1206 NSGALTTLAGSTKGTESVTSSQSGDIGTISGTVGVKATESLTQSNKIKATTTGEA 1265  
  
QY 217 MTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEI 251  
DB 1266 NVTSA GTTGGTISGNTVNTANAGDLTVNGAEI 1300

RESULT 3  
US-08-302-832-2  
; Sequence 2, Application US/08302832  
; Patent No. 5603938  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/302,832  
; FILING DATE: 16-SEP-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US pct/us93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berkstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-404  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1536 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-302-832-2  
  
Query Match 7.5%; Score 98; DB 1; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;  
  
QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINSQALHQAALKPRI-----171  
DB 1147 ITTKGTNNATTGNVEITAGTGSILGSISSSGSVTLTAT-EGALAVSNISGNTVTVA 1205  
  
QY 172 ---AIQRPAGNRIVGDTV-SAMQSGVFWGVIYIS-----LIEGLV-----ARIKAERGE 216  
DB 1206 NSGALTTLAGSTKGTESVTSSQSGDIGTISGTVGVKATESLTQSNKIKATTTGEA 1265  
  
QY 217 MTVIATGGVASLFEGATDSIDHFDSDLTIRGLLEI 251  
DB 1266 NVTSA GTTGGTISGNTVNTANAGDLTVNGAEI 1300

RESULT 4  
US-08-530-198-2  
; Sequence 2, Application US/08530198  
; Patent No. 5869065  
; GENERAL INFORMATION:  
; APPLICANT: BARENKAMP, STEPHEN J  
; APPLICANT: ST. GEME III, JOSEPH W  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS  
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,198  
; FILING DATE: 13-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERKSTRESSER, JERRY W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: JWB-1186

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-530-198-2

Query Match 7.5%; Score 98; DB 2; Length 1536;
Best Local Similarity 29.0%; Pred. No. 0.059;
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIAPGINLSMOALHEAAKLPRI-----171
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSGVSITLAT-EGALAVSNISGNTVTVTA 1205

QY 172 ---AIQRPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV---ARIKAERGE 216
Db 1206 NSGALTTLAGSTIKGTSTSSQSDIGGTISGGTVEVKATESLTQSNKIKATTGEA 1265

QY 217 MTVIATGVASLFEAGTDSIDHFDSDLTIRGLLEI 251
Db 1266 NVTSGTGTGGTISGNTVNTANAGDLTVNGAEI 1300

RESULT 5
US-08-469-880-2
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:

; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 7.5%; Score 98; DB 2; Length 1536;
Best Local Similarity 29.0%; Pred. No. 0.059;
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIAPGINLSMOALHEAAKLPRI-----171
Db 1147 ITTKGTGINATTGNVEITAGTGSILGGIESSGVSITLAT-EGALAVSNISGNTVTVTA 1205

QY 172 ---AIQRPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV---ARIKAERGE 216
Db 1206 NSGALTTLAGSTIKGTSTSSQSDIGGTISGGTVEVKATESLTQSNKIKATTGEA 1265

QY 217 MTVIATGVASLFEAGTDSIDHFDSDLTIRGLLEI 251
Db 1266 NVTSGTGTGGTISGNTVNTANAGDLTVNGAEI 1300

RESULT 6
US-08-728-470-2
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
```

US-08-728-470-2

Query Match 7.5%; Score 98; DB 2; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRI-----171  
DB 1147 ITTKGTTNATTGNVEITATQSGILGSISSGVSLLTAT-EGALAVSNISGNTVTVA 1205  
QY 172 ---AIOBPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGEP 216  
DB 1206 NSGALTTLAGSTIKGTESVTTSSQSDIGTISGTVVEVKATESLTTQSNKIKATTGEA 1265  
QY 217 MVVIATGGVASFEGATDSIDHFDSDLTIRGLLEI 251  
DB 1266 NVTSGTGTIGTISGNTVNTANAGDLTVGNGAEI 1300

## RESULT 7

US-08-617-697-2  
Sequence 2, Application US/08617697  
Patent No. 5977336  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-557  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-617-697-2

Query Match 7.5%; Score 98; DB 2; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRI-----171  
DB 1147 ITTKGTTNATTGNVEITATQSGILGSISSGVSLLTAT-EGALAVSNISGNTVTVA 1205  
QY 172 ---AIOBPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGEP 216  
DB 1206 NSGALTTLAGSTIKGTESVTTSSQSDIGTISGTVVEVKATESLTTQSNKIKATTGEA 1265

DB 1147 ITTKGTTNATTGNVEITATQSGILGSISSGVSLLTAT-EGALAVSNISGNTVTVA 1205  
QY 172 ---AIOBPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGEP 216  
DB 1206 NSGALTTLAGSTIKGTESVTTSSQSDIGTISGTVVEVKATESLTTQSNKIKATTGEA 1265  
QY 217 MVVIATGGVASFEGATDSIDHFDSDLTIRGLLEI 251  
DB 1266 NVTSGTGTIGTISGNTVNTANAGDLTVGNGAEI 1300

## RESULT 8

US-08-719-641-2  
Sequence 2, Application US/08719641  
Patent No. 6218141  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
US-08-719-641-2

Query Match 7.5%; Score 98; DB 4; Length 1536;  
Best Local Similarity 29.0%; Pred. No. 0.059;  
Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;

QY 126 VVIDSGT---ATT--FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRI-----171  
DB 1147 ITTKGTTNATTGNVEITATQSGILGSISSGVSLLTAT-EGALAVSNISGNTVTVA 1205  
QY 172 ---AIOBPAGNRIVGTDV-SAMQSGVFWGYIS-----LIEGLV----ARIKAERGEP 216  
DB 1206 NSGALTTLAGSTIKGTESVTTSSQSDIGTISGTVVEVKATESLTTQSNKIKATTGEA 1265



QY 217 MTVIATGGVASFEGATSDIDFSDLTIRGLLEI 251  
 Db 1266 NVTSGTIGGTISGNTVNTANAGDLTVGNGAEI 1300

## RESULT 9

US-09-206-942-67  
 ; Sequence 67, Application US/09206942  
 ; Patent No. 6432669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H.  
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 ; FILE REFERENCE: 1038-861 MIS:jb  
 ; CURRENT APPLICATION NUMBER: US/09/206,942  
 ; CURRENT FILING DATE: 1998-12-08  
 ; EARLIER APPLICATION NUMBER: 09/167,568  
 ; EARLIER FILING DATE: 1998-10-07  
 ; NUMBER OF SEQ ID NOS: 95  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 67  
 ; LENGTH: 1536  
 ; TYPE: PRX  
 ; ORGANISM: Haemophilus influenzae  
 US-09-206-942-67

Query Match 7.5%; Score 98; DB 4; Length 1536;  
 Best Local Similarity 29.0%; Pred. No. 0.059; Indels 30; Gaps 7;  
 Matches 45; Conservative 22; Mismatches 58; Indels 30; Gaps 7;  
 QY 126 VVIDSGT---APT---FDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRI-----171  
 Db 1147 ITTKGTINATGNVEITAQSGILGGIESSGVTLTAT-EGALAVNSISGNTVTVTA 1205  
 QY 172 ---ATQRPAGNRIVGTDV-SAMQSGVFWGYS-----LIEGLV---ARIKAERGEP 216  
 Db 1206 NSGALTTLAGTSIKGTSTSSQSDIGGTISGGTVKATESLTQTSNSKIKATTGEA 1265  
 QY 217 MTVIATGGVASFEGATSDIDFSDLTIRGLLEI 251  
 Db 1266 NVTSGTIGGTISGNTVNTANAGDLTVGNGAEI 1300

## --RESULT 10

US-09-206-942-28  
 ; Sequence 28, Application US/09206942  
 ; Patent No. 6432669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H.  
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 ; FILE REFERENCE: 1038-861 MIS:jb  
 ; CURRENT APPLICATION NUMBER: US/09/206,942  
 ; CURRENT FILING DATE: 1998-12-08  
 ; EARLIER APPLICATION NUMBER: 09/167,568  
 ; EARLIER FILING DATE: 1998-10-07  
 ; NUMBER OF SEQ ID NOS: 95  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 28  
 ; LENGTH: 1220  
 ; TYPE: PRX  
 ; ORGANISM: Haemophilus influenzae  
 US-09-206-942-28

Query Match 6.9%; Score 90; DB 4; Length 1220;  
 Best Local Similarity 20.3%; Pred. No. 0.35; Indels 66; Gaps 8;  
 Matches 55; Conservative 38; Mismatches 112; Indels 66; Gaps 8;

QY 3 LAIEQGNNTMTFAIHGASWVAOWRSATESSTRTADEYVWVLSQLLSMOGLG-----53  
 Db 135 IAFEKGNNTPTI-----TGOGTITAGNGKGRFFENASLNGIGTGLLFNIKR 179  
 QY 54 -----FRAID-----AVIISVVVPOSIFNRLNLSRRYFNVEPLVIGENAKLGIDV 98  
 Db 180 DLGNFQIINFNGTINISCKVNLSWIPKKNWYSKFRGRTYWNVTHLVNVSSEKFNLT 239  
 QY 99 RIEKPEAGADRLVNAIGAAWYVPGPLVVIDSGTATTFDIVAADG-----APEGG 148  
 Db 240 DSRGDDTAGTLPNTPNLNGISFNKDTTFDVKQNGAVTFDIKAPIGVNNRNLNVSFNGN 299  
 QY 149 I-IAPGINLSMOAL-HEAAKLPRIATQRPAGNRIVGTDVVSAMQSGVFWGYSLIEGLV 206  
 Db 300 ISVGGGNTVTFKLLASSSTAQTGPFVINSKHFNASGGSLEFFRTEGSTKVGFLI-----353  
 QY 207 ARIKAERGEPMTVIATGGVASLFEATDSID 237  
 Db 354 -----NNDLTNATGGNISLQ--VEGID 375

## RESULT 11

US-09-206-942-26  
 ; Sequence 26, Application US/09206942  
 ; Patent No. 6432669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H.  
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 ; FILE REFERENCE: 1038-861 MIS:jb  
 ; CURRENT APPLICATION NUMBER: US/09/206,942  
 ; CURRENT FILING DATE: 1998-12-08  
 ; EARLIER APPLICATION NUMBER: 09/167,568  
 ; EARLIER FILING DATE: 1998-10-07  
 ; NUMBER OF SEQ ID NOS: 95  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 26  
 ; LENGTH: 1226  
 ; TYPE: PRX  
 ; ORGANISM: Haemophilus influenzae  
 US-09-206-942-26

Query Match 6.9%; Score 90; DB 4; Length 1226;  
 Best Local Similarity 20.3%; Pred. No. 0.35; Indels 66; Gaps 8;  
 Matches 55; Conservative 38; Mismatches 112; Indels 66; Gaps 8;

QY 3 LAIEQGNNTMTFAIHGASWVAOWRSATESSTRTADEYVWVLSQLLSMOGLG-----53  
 Db 141 IAFEKGNNTPTI-----TGOGTITAGNGKGRFFENASLNGIGTGLLFNIKR 185  
 QY 54 -----FRAID-----AVIISVVVPOSIFNRLNLSRRYFNVEPLVIGENAKLGIDV 98  
 Db 136 DLGNFQIINFNGTINISCKVNLSWIPKKNWYSKFRGRTYWNVTHLVNVSSEKFNLT 245  
 QY 99 RIEKPEAGADRLVNAIGAAWYVPGPLVVIDSGTATTFDIVAADG-----APEGG 148  
 Db 246 DSRGDDTAGTLPNTPNLNGISFNKDTTFDVKQNGAVTFDIKAPIGVNNRNLNVSFNGN 305  
 QY 149 I-IAPGINLSMOAL-HEAAKLPRIATQRPAGNRIVGTDVVSAMQSGVFWGYSLIEGLV 206  
 Db 306 ISVGGGNTVTFKLLASSSTAQTGPFVINSKHFNASGGSLEFFRTEGSTKVGFLI-----359  
 QY 207 ARIKAERGEPMTVIATGGVASLFEATDSID 237  
 Db 360 -----NNDLTNATGGNISLQ--VEGID 381

## RESULT 12

US-08-804-227C-12  
 ; Sequence 12, Application US/08804227C  
 ; Patent No. 5876991

GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuntz, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-227C-12

Query Match 6.9%; Score 89.5; DB 2; Length 1891;  
Best Local Similarity 27.7%; Pred. No. 0.8;  
Matches 41; Conservative 18; Mismatches 62; Indels 27; Gaps 6;

QY 99 RIEKPEAG-ADRLVNAIGAAMVPGPLVVVDSGTATFDIVAADGAFEGGIIAGINLS 157  
DB 1073 RTETPDRTGLAARLAELARSPEGLAGVLLPLPSGGAA----VAGHPGLDQGTAA--VLLT 1126  
QY 158 MQALHEAAAKLPRIAIORPA-----GNRIVGDTVTSAMQSGVFW-GYISLIEG 204  
DB 1127 IQALTDAAVRAPLWVTVTRGAVAGVSGEVPVCAVGARVWGLGRVAALVPPVQWGLVDVAVG 1186  
QY 205 L-----VARIKAEGERPMTVIATGGV 225  
DB 1187 AGVREWRVVGAVGGEDQVAVRGGV 1214

RESULT 13  
US-08-804-198-6  
Sequence 6, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuntz, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rostock, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/804,198  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1891 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-804-198-6

Query Match 6.9%; Score 89.5; DB 2; Length 1891;  
Best Local Similarity 27.7%; Pred. No. 0.8;  
Matches 41; Conservative 18; Mismatches 62; Indels 27; Gaps 6;

QY 99 RIEKPEAG-ADRLVNAIGAAMVPGPLVVVDSGTATFDIVAADGAFEGGIIAGINLS 157  
DB 1073 RTETPDRTGLAARLAELARSPEGLAGVLLPLPSGGAA----VAGHPGLDQGTAA--VLLT 1126  
QY 158 MQALHEAAAKLPRIAIORPA-----GNRIVGDTVTSAMQSGVFW-GYISLIEG 204  
DB 1127 IQALTDAAVRAPLWVTVTRGAVAGVSGEVPVCAVGARVWGLGRVAALVPPVQWGLVDVAVG 1186  
QY 205 L-----VARIKAEGERPMTVIATGGV 225  
DB 1187 AGVREWRVVGAVGGEDQVAVRGGV 1214

RESULT 14  
US-08-804-198-6  
Patent No. 5183745  
APPLICANT: DANCHIN, ANTOINE; GLASER, PHILLIPPE; KRIN, EVELYN;  
BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES  
TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR  
BIOLOGICAL USES  
NUMBER OF SEQUENCES: 13  
CURRENT APPLICATION DATA:  
FILING DATE: 25-OCT-1989  
SEQ ID NO: 2:  
LENGTH: 1489  
US-08-804-198-6

Query Match 6.5%; Score 85; DB 6; Length 1489;  
Best Local Similarity 20.0%; Pred. No. 1.8;  
Matches 55; Conservative 29; Mismatches 113; Indels 78; Gaps 8;

QY 48 SMOGLGFRAI-----DAVIISVVQSFNLSRRYNNVPLVIGENAKLIDVRIE 101  
DB 407 SLDGVSRSFSLGEVSDMAVAEAAELEMTQVLAGARQDAEPFGVSGASAHWQALQG 466  
QY 102 KPSEAGADRLVNAIGAAMVPG-----PLVVVDSGTATFDIVAADGAF- 146  
DB 467 AQVAAARQLVRAI-ALMTQFGRAGSTNTPQEAASLSAAVFLGEASSAVETVSGFPG 525  
QY 147 -----GGIIAPGINLSMOALHEAAAKLPRIAIORPAGNRIVGDTVSA 189  
DB 526 SSRWAGGFGVAGGAMALGGIIAAVAGAGMSLTDA-----PAGKAAAGAEIAL 574  
QY 190 MOSGVFWGYISLIEGLVARIKAEGERPMTVIATGGVAVSLFEGA----- 232

Db 575 QLTG---GTVELASSIALAARGVTSGLQVAGASAGAAAGALAAALSPMEIYGLVQOS 631

QY 233 ---TDSIDHFDSDLTRG-----LLEIYRNTIAE 259

Db 632 HYADQLDKLAQESSAYGYEGDALLAQLYRDKTAAE 666

## RESULT 15

US-08-669-785-2

; Sequence 2, Application US/08669785

; Patent No. 6309648

; GENERAL INFORMATION:

; APPLICANT: Betsou, Fotini

; APPLICANT: Sebo, Peter

; APPLICANT: Guiso, Nicole

; TITLE OF INVENTION: Protective Epitopes Of Adenyl

; TITLE OF INVENTION: Cyclase-Haemolysin(AC-Hly), Their Application To

; TITLE OF INVENTION: The Treatment Or To The Prevention Of Bordetella Infections.

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &amp; Dunner

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,785

; FILING DATE: 27-JUN-1996

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 02356.0072-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1706 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

US-08-669-785-2

Query Match 6.5%; Score 85; DB 4; Length 1706;

Best Local Similarity 20.0%; Pred. No. 2.3;

Matches 55; Conservative 29; Mismatches 113; Indels 78; Gaps 8;

QY 48 SMOGLGFRAI-----DAVISSVVPQSIENLRNLSRRYFNVEPLVIGENAKLGIDVRIE 101

Db 406 SLDGVGSRFSLSGEVSDMAAEELENTQVLHAGARODDAEPGVSGASAHWGQALQG 465

QY 102 KPSEAGADRLVNAIGAANYPG-----PLVVDSGTATTFDIYADGAFE- 146

Db 466 AQAVAAQRLVHAI-ALMTQFGRAGSTNTPQPAASLSAAVFLGEASSAVAEVTSGFPRG 524

QY 147 -----GGIIAPGINLSMOALHEAAAKLPRTAIORPAGNRIVGTDVSA 189

Db 525 SSRWAGGFGVAGGAMALGGGIAAAGVAGMSLTDDA-----PAGQKAAAGAEIAL 573

QY 190 MQSGVFWGYISLIEGLVARIKAEERGEPTVVIATGGVASLFECA----- 232

Db 574 QLTG---GTVELASSIALAARGVTSGLQVAGASAGAAAGALAAALSPMEIYGLVQOS 630

QY 233 ---TDSIDHFDSDLTRG-----LLEIYRNTIAE 259

Search completed: June 24, 2003, 22:10:43

Job time : 11.5489 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 10.7262 Seconds  
(without alignments)  
2330.267 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEOGNTMFAIHGDA.....SDLTIRGLLEIYRRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	261	2 B87489	transcription acti
2	598	46.0	254	2 F83660	hypothetical prote
3	523	40.3	259	2 AF1102	conserved hypothet
4	519	40.0	273	2 E97293	probable transcrip
5	517	39.8	265	2 T36391	hypothetical prote
6	516	39.7	259	2 AF1464	conserved hypothet
7	492	37.9	233	2 S66100	conserved hypothet
8	396.5	30.5	262	2 E75516	conserved hypothet
9	389.5	30.0	274	2 H86937	conserved hypothet
10	382.5	29.4	272	2 A70955	hypothetical prote
11	359.5	27.7	246	2 D72320	conserved hypothet
12	337.5	26.0	273	2 D71326	conserved hypothet
13	189.5	14.6	262	2 F70165	conserved hypothet
14	175	13.5	592	2 H82031	probable biotin-[a
15	172	13.2	592	2 B81009	Bira protein/Bvg a
16	163.5	12.6	276	2 A12292	hypothetical prote
17	160	12.3	248	2 H83111	hypothetical prote
18	147	11.3	257	2 S75559	hypothetical prote
19	140.5	10.8	242	2 A82637	conserved hypothet
20	134.5	10.4	267	2 I40327	baf protein - Bord
21	125.5	9.7	229	2 E70465	hypothetical prote
22	111	8.5	209	2 H81382	hypothetical prote
23	107.5	8.3	401	2 C83309	conserved hypothet
24	99	7.6	392	2 D70656	hypothetical prote
25	98	7.5	1536	2 A43855	high-molecular-wei
26	97.5	7.5	456	2 JC4089	3alpha,7alpha,12al
27	97.5	7.5	5291	2 F90696	hypothetical prote
28	96.5	7.4	5188	2 B85547	probable RTX fami
29	96	7.4	460	2 AC2272	hypothetical prote

30 95.5 7.4 1682 1 C70588 probable mbtE prot  
31 94.5 7.3 4735 2 T17463 rifamycin polyketi  
32 94 7.2 530 2 AH3194 glycogen synthase  
33 92.5 7.1 657 2 T25613 hypothetical prote  
34 92.5 7.1 701 1 O08CNA carbon starvation  
35 92 7.1 384 2 S75916 probable soluble h  
36 91.5 7.0 386 2 F72773 probable molybdopt  
37 91.5 7.0 487 2 B32475 dnaK-type molecula  
38 91 7.0 327 2 AH2312 glycine cleavage T  
39 91 7.0 464 2 H87208 cystathionine [bet  
40 90.5 7.0 287 2 D90540 glucokinase (gluco  
41 90 6.9 260 2 S36105 hgdC protein - Aci  
42 90 6.9 478 2 T35759 pyruvate kinase -  
43 90 6.9 636 2 AI0057 chaperone protein  
44 89.5 6.9 701 2 E90708 carbon starvation  
45 89.5 6.9 701 2 A85559 carbon starvation

#### ALIGNMENTS

##### RESULT 1

B87489  
transcription activator, probable Baf family [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: B87489  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: B87489  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-261 <STO>  
A:Cross-references: GB:AE005673; NID:gl3423392; PIDN:AAK23910.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1935  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.3lc

Query Match 100.0%; Score 1299; DB 2; Length 261;  
Best Local Similarity 100.0%; Pred. No. 3.7e-96;  
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLLAIEOGNTMFAIHGASVVAQWRSATSTRTADEYVWLSQGLGFRADAV 60  
|||||  
Db 2 MLLAIEOGNTMFAIHGASVVAQWRSATSTRTADEYVWLSQGLGFRADAV 61  
61 IISVVVPSQIFNLRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120  
62 IISVVVPSQIFNLRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 121  
121 YGCLPVVIDSGTATTFDVAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180  
122 YGCLPVVIDSGTATTFDVAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 181  
181 IYGTDTVSAMQSGVFMGYISLIEGLVARIKAEERGEPTVIATGVSFLFEGATDSIDHFD 240  
182 IYGTDTVSAMQSGVFMGYISLIEGLVARIKAEERGEPTVIATGVSFLFEGATDSIDHFD 241  
241 SLDITIRGLLEIYRRNTIAES 260  
242 SLDITIRGLLEIYRRNTIAES 261

##### RESULT 2

F83660  
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: F83660

R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A: Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A: Reference number: A83650; MUID: 20512582; PMID: 11058132  
A: Accession: F83660  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-254 <STO>  
A: Cross-references: GB: AP001507; GB: BA000004; NID: g10172612; PIDN: BAB03805.1; GSPDB: GN00  
A: Experimental source: strain C-125  
C: Genetics:  
A: Gene: BH0086  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 46.0%; Score 598; DB 2; Length 254;  
Best Local Similarity 43.9%; Pred. No. 2.3e-40;  
Matches 112; Conservative 61; Mismatches 80; Indels 2; Gaps 1;

QY 1 MLLATEOGNTNMFATHDGSVAQWRSATSTRTADEYVWLSQGLGFRDAIDAV 60  
1 MLIVDVGNTNVLGYQDETILHWRLATRSQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60  
61 IISVVPOSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIKPESEAGADRLVNAIGAAMV 120  
61 VISSVPPMFSLEOMCKKYFHTPMIGPGIKTGLNIDYDNPKEVGADRIYVNAIAIEL 120  
QY 121 YGPLVVIDSGTATTFDIIAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180  
121 YGPAIVVDFGATTVCLINEKQVAGGVIAFGIMISTEALYHRASKLPRIETAKP--KQ 178  
QY 181 IYGTDTVSAMQSGVFWGYISLIEGLVARIKAEKPEMTVIATGGVASLFEATDSIDHFD 240  
179 VVGTIDISMQSGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVID 238  
QY 241 SLDITIRGLLEIYRN 255  
239 SFLTLKGLLIYKKN 253

RESULT 3  
AF1102  
conserved hypothetical protein lmo0221 [imported] - *Listeria monocytogenes* (strain EGD-e  
C: Species: *Listeria monocytogenes*  
C: Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C: Accession: AF1102  
R: Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A: Title: Comparative genomics of *Listeria* species.  
A: Reference number: AB1077; MUID: 21537279; PMID: 11679669  
A: Accession: AF1102  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-259 <GLA>  
A: Cross-references: GB: NC\_003210; PIDN: CAD00748.1; PID: g16409586; GSPDB: GN00177  
A: Experimental source: strain EGD-e  
C: Genetics:  
A: Gene: lmo0221  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 40.3%; Score 523; DB 2; Length 259;  
Best Local Similarity 39.2%; Pred. No. 2.3e-34;  
Matches 100; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

QY 1 MLLATEOGNTNMFATHDGSVAQWRSATSTRTADEYVWLSQGLGFRDAIDAV 60  
1 MLIVDVGNTNVLGYQDETILHWRLATRSQKTEDEYAMTVRSFLDHAGLQFQDIDGI 60  
61 IISVVPOSIFNLNLSRRYFNVEPLVIGENAKLGIDVRIKPESEAGADRLVNAIGAAMV 120  
61 VISSVPPMFSLEOMCKKYFHTPMIGPGIKTGLNIDYDNPKEVGADRIYVNAIAIEL 120  
QY 121 YGPLVVIDSGTATTFDIIAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180  
121 YGPAIVVDFGATTVCLINEKQVAGGVIAFGIMISTEALYHRASKLPRIETAKP--KQ 178  
QY 181 IYGTDTVSAMQSGVFWGYISLIEGLVARIKAEKPEMTVIATGGVASLFEATDSIDHFD 240  
179 VVGTIDISMQSGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVID 238  
QY 241 SLDITIRGLLEIYRN 255  
239 SFLTLKGLLIYKKN 253

RESULT 5  
T36391  
hypothetical protein SCE94.31c - *Streptomyces coelicolor*  
C: Species: *Streptomyces coelicolor*  
C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
C: Accession: T36391  
R: Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.;  
submitted to the EMBL Data Library, April 1999  
A: Reference number: Z21573  
A: Accession: T36391  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA

Db 61 IISVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNKLVDNPREIGSDRIVNAVAASEE 120  
QY 121 YGPLVVIDSGTATTFDIIAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180  
Db 121 YGTPVIVDFGATTVCLINEKQVAGGVIAFGIMISTEALYHRASKLPRIETAKP--SSQ 178  
QY 181 IYGTDTVSAMQSGVFWGYISLIEGLVARIKAEKPEMTVIATGGVASLFEATDSIDHFD 240  
Db 179 IIGKSVSSMQAGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVID 238  
QY 241 SLDITIRGLLEIYRN 255  
Db 239 SFLTLKGLLEIYRN 253

RESULT 4  
E97293  
probable transcription regulator, homolog of Bvg accessory factor [imported] - *Clostr*  
C: Species: *Clostridium acetobutylicum*  
C: Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C: Accession: E97293  
R: Noll, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;  
Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A: Title: Genome Sequence and Comparative Analysis of the Solvent-Producing *Bacterium*  
A: Reference number: A96900; MUID: 21359325; PMID: 21359325  
A: Accession: E97293  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-273 <KUR>  
A: Cross-references: GB: AE001437; PIDN: AAK81136.1; PID: g15026270; GSPDB: GN00168  
A: Experimental source: *Clostridium acetobutylicum* ATCC824  
C: Genetics:  
A: Gene: CAC3200  
C: Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 40.0%; Score 519; DB 2; Length 273;  
Best Local Similarity 40.1%; Pred. No. 5e-34;  
Matches 105; Conservative 58; Mismatches 91; Indels 8; Gaps 3;

QY 1 MLLATEOGNTNMFATHDGSVAQWRSATSTRTADEYVWLSQGLGFRDAIDAV 60  
121 YGPLVVIDSGTATTFDIIAADGAFEGGIAPGINLSMQALHEAAKLPRIATQRPAGNR 180  
121 YGTPVIVDFGATTVCLINEKQVAGGVIAFGIMISTEALYHRASKLPRIETAKP--SSQ 178  
QY 181 IYGTDTVSAMQSGVFWGYISLIEGLVARIKAEKPEMTVIATGGVASLFEATDSIDHFD 240  
179 IIGKSVSSMQAGIFGYVSDGVVKKRKAQAESEPKVIATGGLAKLIGTSETIDVID 238  
QY 241 SLDITIRGLLEIYRN 255  
239 SFLTLKGLLEIYRN 253

RESULT 5  
T36391  
hypothetical protein SCE94.31c - *Streptomyces coelicolor*  
C: Species: *Streptomyces coelicolor*  
C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jun-2000  
C: Accession: T36391  
R: Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.;  
submitted to the EMBL Data Library, April 1999  
A: Reference number: Z21573  
A: Accession: T36391  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA

\* A:Residues: 1-265 <OLI>  
A:Cross-references: EMBL:AL049628; PIDN: CAB40880.1; GSPDB: GN00070; SCODB: SCE94.31c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCODB: SCE94.31c  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 39.8%; Score 517; DB 2; Length 265;  
Best Local Similarity 42.4%; Pred. No. 7e-34;  
Matches 112; Conservative 50; Mismatches 90; Indels 12; Gaps 5;  
Qy 1 MLLAEQNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMQLSGFRDAV 55  
Db 1 MLLIDVGNTHVLGDFGDIVHEHRISTDSRTADELAVLQGLMGHPLLDGLG-D 59  
Qy 56 AIDAVIISVVQSFNRLNLSRRYFNVEPLVIGE-NAKLGIDVRIEKPSEAGADRLVNA 114  
Db 60 GIDGIAICATVPSVLHRELRETRRYGDPVAVLPGVKTGVPILTDHPKEVGADRIINA 119  
/ 115 IGAAWVYGPVLVWIDSGTATTFDVAADGAFEGGIIAPGINSQALHEAAKLPRIAIQ 174  
Db 120 VAAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAFGIISVEALGVKGAQLRKIEVA 179  
Qy 175 RPAGNRIVGTIVSAMQSGVFWGYISLIEGLV--ARIKAERGEPTMTVIATGGVSLFEG 231  
Db 180 RP--RSVIGKNTVEMQSGIYVGFAGQVDGVNRMARELADDDVTVIATGGLAPMWIG 237  
Qy 232 ATDSIDHPSDLTIRGLLEIYRN 255  
Db 238 ESSVIDEHEPWLTMLGLRLVYRN 261

RESULT 6  
AF1464  
conserved hypothetical protein lin0253 [Imported] - Listeria innocua (strain Clip11262)  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1464  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Meakins, C.; Schlueper, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID: 21537279; PMID: 11679669  
A:Accession: AF1464  
Status: preliminary  
Molecule type: DNA  
A:Residues: 1-259 <GLA>  
A:Cross-references: GB:AL592022; PIDN: CAC95486.1; PID: g16412682; GSPDB: GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: lin0253  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 39.7%; Score 516; DB 2; Length 259;  
Best Local Similarity 39.2%; Pred. No. 8.1e-34;  
Matches 100; Conservative 58; Mismatches 95; Indels 2; Gaps 1;  
Qy 1 MLLAEQNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMQLSGFRDAV 60  
Db 1 MLVIDVGNTHVLTGVYHDKLEYHWRIFTSRHKTEDEFGMILSLFDSHGLMFEQIDGI 60  
Qy 61 IISVVPOSIFNRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGRAV 120  
Db 61 IISVVPPIMHAMTCVFNIRPLIVGPIKTLGNLKVNDPREIGSDIRVNAVAASEE 120  
Qy 121 YPGPLVWIDSGTATTFDVAADGAFEGGIIAPGINSQALHEAAKLPRIAIORPAGNR 180  
Db 121 YGTPVIVVDGTATTFYIDGAVYQGGATAPGIMNISTEALYNRAKLPRIADAE--SSQ 178  
Qy 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIKAEERGEPTMTVIATGGVSLFEGATDSIDHFD 240

Db 179 IIGKSTVSMQAGIFYGIGQCEGIIAEMKQSNTPVYVATGGLARMITEKSSAYDILD 238  
Qy 241-SDLTIRGLLEIYRN 255  
Db 239 PFLTUKGLELLIYRN 253  
RESULT 7  
S66100  
conserved hypothetical protein yacB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 18-Aug-2000  
C:Accession: S66100; E69740  
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994  
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis  
A:Reference number: S65967; MUID: 96051385; PMID: 7584024  
A:Accession: S66100  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <OGA>  
A:Cross-references: EMBL:D26185; NID: 9467326; PIDN: BAA05305.1; PID: d1005847; PID: 9467326  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1994  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Bairoch, A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Giesch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardir, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meyer, M.; Ogasawara, N.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scar, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Skeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchly, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshikawa, T.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis  
A:Reference number: A69580; MUID: 98044033; PMID: 9384377  
A:Accession: E69740  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <KUN>  
A:Cross-references: GB:Z99104; GB:AL009126; NID: 92632267; PIDN: CAB11846.1; PID: e1182  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yacB  
A:Start codon: TTG  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.9%; Score 492; DB 2; Length 233;  
Best Local Similarity 45.2%; Pred. No. 5.8e-32;  
Matches 95; Conservative 41; Mismatches 72; Indels 2; Gaps 1;  
Qy 1 MLLAEQNTNTMFAIHGASWVAQWRSATSTRTADEYVWVLSQLLSMQLSGFRDAV 60  
Db 1 MLLVIDVGNTHVLTGVYHDKLEYHWRIFTSRHKTEDEFGMILSLFDSHGLMFEQIDGI 60  
Qy 61 IISVVPOSIFNRLNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGRAV 120  
Db 61 IISVVPPIMFALERMCTKYHFIEPQIVGPGMKTLGNIKIDYFNKPKVEGADRIVNAVAAILH 120  
Qy 121 YPGPLVWIDSGTATTFDVAADGAFEGGIIAPGINSQALHEAAKLPRIAIORPAGNR 180  
Db 121 YGNPLIVVDGTATTFYIDENKQYMGGAIPGISTEALYSRAKLPRIETRP--DN 178  
Qy 181 IVGTDVTSAMQSGVFWGYISLIEGLVARIK 210  
Db 179 IIGKNTVSMQSGILFGVGVQVEGIVKRMK 208

RESULT 8

E75516  
conserved hypothetical protein - Deinococcus radiodurans (strain RL)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Jun-2000  
C:Accession: E75516  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: E75516  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <WHI>  
A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g645814  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0461  
A:Map position: 1  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.5%; Score 396.5; DB 2; Length 262;  
Best Local Similarity 36.1%; Pred. No. 2.7e-24; Mismatches 9; Gaps 5;  
Matches 92; Conservative 45; Indels 109; Indels 9; Gaps 5;  
QY 2 LLAIEQGNNTMFAIHGASVAQ-WRSATESRTADEYVWVLSQLLSMOGLGFRDAIDAV 60  
DB 6 LLAVIDGNNTTVGLADASGALHTWRTNREMLPDDLALQLHGLFTLAGAPR-AA 63  
QY 61 IISVVVPSQIFNLRLNLSRYFVNEPLVIGENAKLGIDVRIKPSAGADRLYNNAI 120  
DB 64 VLSSVAPPVGENYALAKRHFMDAFVSAENLPDVTVELDTPGSGADRLCNLFGEAK- 122  
QY 121 YEGPL---VVIDSGTATFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIATORPA 177  
DB 123 YLGGLDYAVVDFGTSTFEDVGRGRFGLGILATCAQVSADALFAAKLPRIITQAP- 181  
QY 178 GNRIVGTDTVSAMQSGVFWYISLIEGLVARIKAEGERPMTVIATGGVASLFEATDSD 237  
DB 182 -ETAIGNVTHALQSLVFGYAEVMDGLLRIRAEPLGGEAVAVATGGFSTVGICQCID 240  
QY 238 HFDSDLTIRGLLEIY 252  
DB 241 YYDETTLRGLVELW 255

## RESULT 9

H86937  
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae  
Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: H86937  
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; MUID:21128732; PMID:11234002  
A:Accession: H86937  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-274 <STO>  
A:Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: ML0232  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.0%; Score 389.5; DB 2; Length 274;  
Best Local Similarity 34.3%; Pred. No. 1e-23;  
Matches 93; Conservative 53; Mismatches 108; Indels 17; Gaps 5;

QY 1 MLLAIEQGNNTMFAIHG-----ASWVAQWRSATESRTADEYVWVLSQLLSMOGLGFRA 56  
DB 1 MLLAIDVRNTHTVVGLSGSKEHAKVYQWRIRTESEVTADALALIIDGLI---GDSER 57  
QY 57 IDAVIISVVPSQIFNLRLNLSRYF-NVEPLVIGENAKLGIDVRIKPSAGADRLYNNAI 115  
DB 58 LAGAAALSTVPSVHVEVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVAGADRIYNCL 117  
QY 116 GAAMVYPGPLVVIDSGTATFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIATOR 175  
DB 118 AAFHKEGQAIIIVDFGSSICVDVVSAGFELGGAIAFGVQVSSDAAAARSAAALRRVELAR 177  
QY 176 PAGNRIVGTDTVSAMQSGVFWYISLIEGLVARIK-----AERGPMVTIATGGVASL 228  
DB 178 P--RSVVGKNTVECMQAGVFGFAGLVGLVGRMQDVEFGDGLGNRVAVVATGHTAPL 235  
QY 229 FEGATSDIHFDSDLTIRGLLEIYRNTIAE 259  
DB 236 LLPELHTVDHYDRHLTLHGLRLVFERNREAO 266

## RESULT 10

A70955  
hypothetical protein RV3600c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: A70955  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70955  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-272 <COL>  
A:Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3600c  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 29.4%; Score 382.5; DB 2; Length 272;  
Best Local Similarity 34.0%; Pred. No. 3.7e-23;  
Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY 1 MLLAIEQGNNTMFAIHG-----ASWVAQWRSATESRTADEYVWVLSQLLSMOGLGFRA 56  
DB 1 MLLAIDVRNTHTVVGLSGSKEHAKVYQWRIRTESEVTADALALIIDGLI---GDSER 57  
QY 57 IDAVIISVVPSQIFNLRLNLSRYF-NVEPLVIGENAKLGIDVRIKPSAGADRLYNNAI 115  
DB 58 LAGTAAALSTVPSVHVEVRIMLDQWPSVPHVLEPGVGTGIPLLVDNPKVAGADRIYNCL 117  
QY 116 GAAMVYPGPLVVIDSGTATFDIVAADGAFEGGIAPGINLSMQALHEAAKLPRIATOR 175  
DB 118 AAYDRFKAIIIVDFGSSICVDVVSAGFELGGAIAFGVQVSSDAAAARSAAALRRVELAR 177  
QY 176 PAGNRIVGTDTVSAMQSGVFWYISLIEGLVARIKAEGERPMTVIATGGVASLFE 230  
DB 178 P--RSVVGKNTVECMQAGVFGFAGLVGLVGRMQDVEFGDGLGNRVAVVATGHTAPL 235  
QY 231 GATDSIDHFDSDLTIRGLLEIYRNN 255  
DB 236 PELHTVDHYDQHLTLQGLRLVFERN 260

## RESULT 11

D72320  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima



C; Superfamily: Streptomyces coelicolor hypothetical protein SCE94\_31c

RESULT 13  
F70165  
conserved hypothetical protein BB0527 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: F70165  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra,  
son, D.; Peterson, J.; Kervavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Ha  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: F70165  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <KUE>  
A:Cross-references: GB:AF001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; P  
A:Experimental source: strain B31

```

Db      9 LIIDIGNTAFALFKDNQVNLFKTKMTNLMLRYDEVYFFENEDFN-----VKNVFI 62
QY      63 SSVVPOQISFNLNLSRRYFNVEPIVIGENAKLIGIDVRIEKPE--AGADRLVNAIGAAMV 120
Db      63 SSVVPIILNEFKNVIFSEFKIKPLFEGFDLNVDTFNPKSKDFLLGSDVFANLVAIEN 122
QY      121 YP-GPLVVDSGTATTTDVIANDGAFEGGIIAPGILNSQWALHEAAKLPRIAIORPAGN 179
Db      123 YSFENVLVLDLGTACTIAVFSRQDQILGILGINSGLPINFNSLUDNAYLLKKFPITSP--N 180
QY      180 RYGTGTVSAMOSQVFWGYYIILBGLVARIKAERGEPMTVIATGVASLFEGATSDSIDHF 239
Db      181 NLLERTSGSVNSGLFYQYKYLEGVYEDIKQMYKKKNLITGTGNADLLSLIETEFIF 240
QY      240 DSDLTIRGL 248
Db      241 NIHLTVEGV 249

```

RESULT 14  
H82031  
Probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis  
N; Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)  
C; Species: Neisseria meningitidis  
C; Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 03-Jun-2002  
C; Accession: H82031  
K; Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Riddle, G.S.; Whitehead, P.M.A.; Brown, A.R.; Davis, A.P.; Dougan, G.; Fraser, H.C.; Harris, D.S.; Holt, K.E.; Jones, A.; Karp, P.; Linton, L.; Martin, J.F.; Morgan, T.; Parkhill, J.; Paulsen, O.; Pearson, T.A.; Saunders, N.J.; Seeger, A.P.; Skrzypek, F.S.; Staden, R.; Struelens, L.; Taggart, J.B.; Travis, J.; Unwin, P.L.; Venter, A.; Whitehead, P.M.A.; Woodhouse, M.; Young, I.P.; Barrett, T.J.; Berridge, M.; Brooks, K.; Brown, J.H.; Brown, S.; Brown, W.; Brown, Z.; Burrows, W.; Carver, T.F.; Chillingworth, T.; Clark, F.; Clifton, E.; Collins, M.; Connor, T.; Coulson, A.; Davies, K.; Deeks, J.G.; Denby, L.; Drenth, J.; Durrant, J.; Earls, J.; Ellis, S.; Ellis, W.; Evans, S.; Evans, T.; Feeney, A.; Field, S.; Fields, C.; Foster, J.; Fraser, J.C.; Freeman, B.; Freeman, C.; Fry, B.; Fry, S.; Fry, T.; Fry, W.; Fry, Y.; Fry, Z.; Fry, A.; Fry, B.; Fry, C.; Fry, D.; Fry, E.; Fry, F.; Fry, G.; Fry, H.; Fry, I.; Fry, J.; Fry, K.; Fry, L.; Fry, M.; Fry, N.; Fry, O.; Fry, P.; Fry, Q.; Fry, R.; Fry, S.; Fry, T.; Fry, U.; Fry, V.; Fry, W.; Fry, X.; Fry, Y.; Fry, Z.; Fry, A.; Fry, B.; Fry, C.; Fry, D.; Fry, E.; Fry, F.; Fry, G.; Fry, H.; Fry, I.; Fry, J.; Fry, K.; Fry, L.; Fry, M.; Fry, N.; Fry, O.; Fry, P.; Fry, Q.; Fry, R.; Fry, S.; Fry, T.; Fry, U.; Fry, V.; Fry, W.; Fry, X.; Fry, Y.; Fry, Z.; Fry, A.; Fry, B.; Fry, C.; Fry, D.; Fry, E.; Fry, F.; Fry, G.; Fry, H.; Fry, I.; Fry, J.; Fry, K.; Fry, L.; Fry, M.; Fry, N.; Fry, O.; Fry, P.; Fry, Q.; Fry, R.; Fry, S.; Fry, T.; Fry, U.; Fry, V.; Fry, W.; Fry, X.; Fry, Y.; Fry, Z;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

-OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 5.23014 Seconds  
(without alignments)  
2061.866 Million cell updates/sec

Title: US-09-813-453a-51

Perfect score: 1299

Sequence: 1 MLLALEGQNTNFPAIHGDA.....SDLTIRGLLEIYRNTAIES 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

~earched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	492	37.9	233	1 YACB_BACSU	P37564 bacillus su
2	134.5	10.4	267	1 BAF_BORPE	Q45338 bordetella
3	94	7.2	496	1 NIFE_RHIL0	Q98ap4 rhizobium l
4	93	7.2	491	1 YC70_CORGL	P42531 corynebacte
5	92.5	7.1	657	1 H57F_CAEEL	P11141 caenorhabdi
6	92.5	7.1	701	1 C5TA_ECOLI	P15078 escherichia
7	92	7.1	308	1 SPD1_DATST	Q96556 datura stra
8	90	6.9	260	1 HGDC_ACIFE	P11568 acidaminoco
9	90	6.9	496	1 NIFE_RHISN	P55673 rhizobium s
10	90	6.9	636	1 DNAK_YERPE	Q82im7 yersinia p
11	89.5	6.9	478	1 CARL_HUMAN	Q9uhj6 homo sapien
12	89	6.9	556	1 CG48_SCHPO	P78750 schizosacch
13	88.5	6.8	469	1 LEU2_RHIL0	Q98ef1 rhizobium l
14	88	6.8	399	1 PKC_RHIME	Q92m79 rhizobium m
15	88	6.8	1307	1 BCCA_ACEXY	Q9wx71 acetobacter
16	87.5	6.7	245	1 H1S4_BACSU	Q35006 bacillus su
17	87.5	6.7	541	1 YQEB_ECOLI	Q46808 escherichia
18	87.5	6.7	555	1 YP85_MYCLE	Q49646 mycobacteri
19	87.5	6.7	576	1 MYTL_CHLMU	Q9p195 chlamydia m
20	87.5	6.7	923	1 GYRA_PSEAE	P48372 pseudomonas
21	87	6.7	201	1 RISA_MYCTU	P1680 mycobacteri
22	87	6.7	637	1 DNAK_SALTI	Q82r1 salmonella
23	87	6.7	637	1 DNAK_SALTY	Q56073 salmonella
24	86.5	6.7	343	1 YJCR_ECOLI	P32716 escherichia
25	86.5	6.7	548	1 YAVL_RHISN	Q53217 rhizobium s
26	86.5	6.7	1132	1 PHYL_PHPYA	P36505 physcomitre
27	86	6.6	537	1 CH61_MYCLE	P37508 mycobacteri
28	86	6.6	788	1 PUR2_YARLI	Q99148 y bifunctio
29	85	6.5	637	1 DNAK_ECOLI	P04475 escherichia
30	85	6.5	1706	1 CYAA_BORPE	P15318 bordetella
31	84.5	6.5	1705	1 CYAA_BORBR	Q57506 bordetella
32	83	6.4	465	1 LEU2_ECO57	Q8xa00 escherichia
33	83	6.4	465	1 LEU2_ECOLI	P30127 escherichia

34	83	6.4	541	1 LCPT_BACSU	P55910 bacillus su
35	82.5	6.4	286	1 PTND_ECOLI	P08188 escherichia
36	82.5	6.4	469	1 LEU2_AGR75	Q8ub99 agrobacteri
37	82.5	6.4	476	1 LEU2_YERPE	Q82ih0 yersinia pe
38	82	6.3	609	1 COPA_PSESM	P12374 pseudomonas
39	81.5	6.3	434	1 ENO_STRPY	P82479 streptococc
40	81.5	6.3	769	1 VPI_BPPH6	P11126 bacteriopho
41	81	6.2	363	1 MODC_RHOCA	Q08381 rhodobacter
42	81	6.2	487	1 IMDH_PASMO	Q916b7 pasteurella
43	81	6.2	637	1 DNAK_BRUME	Q08y76 brucella me
44	81	6.2	644	1 DNAK_NITEU	Q06430 nitrosomona
45	81	6.2	654	1 DNAK_CHLMU	P56836 chlamydia m

#### ALIGNMENTS

RESULT 1  
YACB\_BACSU  
ID YACB\_BACSU STANDARD; PRT; 233 AA.  
AC P37564;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein yacB.  
GN YACB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Bourissier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.N., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Noone D., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivolta C., Rocha E., Roche R., Rose M., Sadale Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
subtilis.";  
RL Nature 390:249-256(1997).  
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D26185; BAA05305.1; -  
DR EMBL; 299104; CAB11846.1; -  
DR SubtilList; BG10133; YACB.  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRFAMs; TIGR00671; baf; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 233 AA; 26217 MW; AAE96E732C15DF44 CRC64;  
  
Query Match 37.9%; Score 492; DB 1; Length 233;  
Best Local Similarity 45.2%; Pred. No. 2.3e-31;  
Matches 95; Conservative 41; Mismatches 72; Indels 2; Gaps 1;  
  
-Y 1 MLLATEQNTNMFATHDGCASVAOWRSATSTRTADEYVWLSQLLSMQGLGFAIDAV 60  
Db 1 MLLYIDVGNNTVGLVYHDKLEYHWRIETSHKTEDEFGMLRSLEDSGLMFEQIDGI 60  
QY 61 IISVVVPOSIFNLRLSRFYVNEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMY 120  
Db 61 IISVVVPIINFALERMCTFYHIEPQIVGPKMTGLNIDNPKEVGADRIVNAVAIHL 120  
QY 121 YGCLVVIDSGTATFDIADGAFEGGIIAPGINSQALHEAAKLPRIAIQRPAGNR 180  
Db 121 YGNPLIVDFGATTTCYIDENQYMGGAIFCITISTEALYSRAKLPRIETTRP--DN 178  
QY 181 IVGTDTVSAMQSGVFWGYSILIEGLVARIK 210  
Db 179 IIGKNTVSAMQSGILFYGVQGVGIVKRMK 208  
  
RESULT 2  
ID BAF\_BORPE STANDARD; PRT; 267 AA.  
AC Q45338; Q45373;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Bvg accessory factor.  
GN BAF.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BP504;  
RX MEDLINE=95325323; PubMed=7601846;  
RA Deshazer D., Wood G.E., Friedman R.L.;  
RT "Identification of a Bordetella pertussis regulatory factor required  
RT for transcriptions of the pertussis toxin operon in Escherichia  
RT coli.";  
RL J. Bacteriol. 177:3801-3807(1995).  
RN [2]  
RP SEQUENCE OF 1-38 FROM N.A.  
RC STRAIN-BP504;  
RX MEDLINE=96419162; PubMed=8821935;  
RA Allen A.G., Maskell D.J.;  
RT "The identification, cloning and mutagenesis of a genetic locus  
RT

required for lipopolysaccharide biosynthesis in Bordetella  
RT pertussis.";  
RL Microbiol. 19:37-52(1996).  
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN  
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF  
CC RNA POLYMERASE.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U12020; AAA75361.1; -  
DR EMBL; AF016461; AAC68834.1; -  
DR EMBL; X90711; CAA62242.1; -  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
KW Transcription regulation; Activator.  
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;  
  
Query Match 10.4%; Score 134.5; DB 1; Length 267;  
Best Local Similarity 28.8%; Pred. No. 0.0014;  
Matches 53; Conservative 18; Mismatches 78; Indels 35; Gaps 8;  
  
QY 79 RYFNVPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMYPG---PLVVIDSQTATT 135  
Db 81 RWLRAQPLA-----MGLRNGYRNPDLGADRACWGVLAROPSVHPPLLVASFQTATT 134  
QY 136 FTIVAADGAFEGGIIAPGINSQALHEAAKLP---RIAQRPAQNRIVGTDTVSAMQS 192  
Db 135 LDTIGDNPVPGGLIPGPAMRGALAYGTGHTPLADGLVADYPIDTHQAIAAGAAQA 194  
QY 193 GVF---W-----GYISLIEGLVA-----RIKAERCEPMTVATGGVASLFEQATSID 237  
Db 195 GAIVRWLAGRORYGOAPEIYVAGGWPVQRAER-----LLAVTGAAF---GATPOPT 246  
QY 238 HFDS 241  
Db 247 YLDS 250  
  
RESULT 3  
ID NIFE\_RHILO STANDARD; PRT; 496 AA.  
AC Q98AP4;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Nitrogenase iron-molybdenum cofactor biosynthesis protein nife.  
GN NIFE OR MLE5908.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF.  
CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).  
CC -!- PATHWAY: Fe-Mo cofactor biosynthesis.  
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U88315; AAB42371.1; -  
 DR EMBL; X07678; CAA30525.1; -  
 DR PIR; S03210; S03210.  
 DR PIR; B32475; B32475.  
 DR HSP; P04475; LDG4.  
 DR WormPep; C37H5.8; CE08631.  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam; PF00012; Hsp70; 1.  
 DR PRINTS; PRO0301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; Hsp70.1; 1.  
 DR PROSITE; PS00329; Hsp70.2; 1.  
 DR PROSITE; PS01036; Hsp70.3; 1.  
 KW ATP-binding; Heat shock; Multigene family; Mitochondrion;  
 KW Transit peptide.  
 FT TRANSIT 1 27 MITOCHONDRION (POTENTIAL).  
 CHAIN 28 657  
 CONFLICT 138 138  
 CONFLICT 140 140 V -> E (IN REF. 2).  
 CONFLICT 167 167 S -> P (IN REF. 2).  
 CONFLICT 347 348 A -> P (IN REF. 2).  
 CONFLICT 347 348 KA -> NV (IN REF. 2).  
 SQ SEQUENCE 657 AA; 70844 MW; 637A1636B6AB836 CRC64;  
 Query Match 7.1%; Score 92.5; DB 1; Length 657;  
 Best Local Similarity 21.9%; Pred. No. 7.1;  
 Matches 57; Conservative 44; Mismatches 102; Indels 57; Gaps 13;  
 QY 2 LLAIEQGNNTWFAHDSASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRADAVI 61  
 DB 33 VIGIDLGTTSCVSTMEGKT--PKVIENAGVTRTPTSTVAFTAD-----GERLVGAP 83  
 QY 62 ISSVVPQSFNL-----RNLRRYFNV-----PLVIGENAKLGIDV--RIEKPS 105  
 DB 84 KRQAVTNSANTLFATKRLIGRIYDEPEVQKDLKVPYKIVKASNGDAWVEAQGVYSPSQ 143  
 QY 106 AGADLVNAIGAAMVYVPGVVIDSGTATT---FDVAADGAFEGGIIAPGINSMOALH 162  
 DB 144 VGFVLMKMKETAESYLG--TTVNNAVTVFAYFNDSSQATKDAQIS-GLNV-LRVIN 199  
 QY 163 EAAKLPRIATQRPAGNRIV-----GTDTSVAM--QSGVF-----WGY 198  
 DB 200 EPTAAALAYGLDKDAGDKIIAYVDLGGGTFDVSILEIQGVPEVKSTNGDTFLGGDFDH 259  
 QY 199 ISLIEGLVARIKAERGEPT 218  
 DB 260 -ALVHLVGEFKEQGVDT 278

## RESULT 6

CC EMBL; U88315; AAB42371.1; -  
 CC EMBL; X07678; CAA30525.1; -  
 CC PIR; S03210; S03210.  
 CC PIR; B32475; B32475.  
 CC HSP; P04475; LDG4.  
 CC WormPep; C37H5.8; CE08631.  
 CC InterPro; IPR001023; Hsp70.  
 CC Pfam; PF00012; Hsp70; 1.  
 CC PRINTS; PRO0301; HEATSHOCK70.  
 CC ProDom; PD000089; Hsp70; 1.  
 CC PROSITE; PS00297; Hsp70.1; 1.  
 CC PROSITE; PS00329; Hsp70.2; 1.  
 CC PROSITE; PS01036; Hsp70.3; 1.  
 KW ATP-binding; Heat shock; Multigene family; Mitochondrion;  
 KW Transit peptide.  
 FT TRANSIT 1 27 MITOCHONDRION (POTENTIAL).  
 CHAIN 28 657  
 CONFLICT 138 138  
 CONFLICT 140 140 V -> E (IN REF. 2).  
 CONFLICT 167 167 S -> P (IN REF. 2).  
 CONFLICT 347 348 A -> P (IN REF. 2).  
 CONFLICT 347 348 KA -> NV (IN REF. 2).  
 SQ SEQUENCE 657 AA; 70844 MW; 637A1636B6AB836 CRC64;  
 Query Match 7.1%; Score 92.5; DB 1; Length 657;  
 Best Local Similarity 21.9%; Pred. No. 7.1;  
 Matches 57; Conservative 44; Mismatches 102; Indels 57; Gaps 13;  
 QY 2 LLAIEQGNNTWFAHDSASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRADAVI 61  
 DB 33 VIGIDLGTTSCVSTMEGKT--PKVIENAGVTRTPTSTVAFTAD-----GERLVGAP 83  
 QY 62 ISSVVPQSFNL-----RNLRRYFNV-----PLVIGENAKLGIDV--RIEKPS 105  
 DB 84 KRQAVTNSANTLFATKRLIGRIYDEPEVQKDLKVPYKIVKASNGDAWVEAQGVYSPSQ 143  
 QY 106 AGADLVNAIGAAMVYVPGVVIDSGTATT---FDVAADGAFEGGIIAPGINSMOALH 162  
 DB 144 VGFVLMKMKETAESYLG--TTVNNAVTVFAYFNDSSQATKDAQIS-GLNV-LRVIN 199  
 QY 163 EAAKLPRIATQRPAGNRIV-----GTDTSVAM--QSGVF-----WGY 198  
 DB 200 EPTAAALAYGLDKDAGDKIIAYVDLGGGTFDVSILEIQGVPEVKSTNGDTFLGGDFDH 259  
 QY 199 ISLIEGLVARIKAERGEPT 218  
 DB 260 -ALVHLVGEFKEQGVDT 278  
 RESULT 6  
 ID CSTA\_ECOLI STANDARD; PRT; 701 AA.  
 AC P15078; P23517; P77740;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Carbon starvation protein A.  
 GN CSTA OR B0598.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-91162636; PubMed-1848300;  
 RA Schultz J.E., Matin A.;  
 RT "Molecular and functional characterization of a carbon starvation  
 RT gene of Escherichia coli."  
 RL J. Mol. Biol. 218:129-140(1991).

RN SEQUENCE FROM N.A.  
 RP STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-200 FROM N.A.  
 RX MEDLINE-89123155; PubMed-2521622;  
 RA Liu J., Duncan K., Walsh C.T.;  
 RT "Nucleotide sequence of a cluster of Escherichia coli enterobactin  
 RT biosynthesis genes: identification of entA and purification of its  
 RT product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase."  
 RL J. Bacteriol. 171:791-798(1989).  
 CC -!- FUNCTION: PEPTIDE UTILIZATION DURING CARBON STARVATION.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -!- INDUCTION: By carbon starvation.  
 CC -!- SIMILARITY: BELONGS TO THE CSTA FAMILY.  
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 CC FRAMESHIFT IN POSITION 543.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

RESULT 8		
HGDC_ACIFE	STANDARD;	PRT; 260 AA.
ID	HGDC_ACIFE	
AC	P11568; Q44042;	
DT	01-OCT-1989 (Rel. 12, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Activator of (R)-2-hydroxyglutaryl-CoA dehydratase (2-hydroxyglutary	
DE	CoA dehydratase component A).	
GN	HGDC.	

RN	[2]	
RP	SEQUENCE OF 129-260 FROM N.A.	
RC	STRAIN=ATCC 25085;	
RX	MEDLINE=89276363; PubMed=2659350;	
RA	Dutscho R., Wohlfarth G., Buckel W.;	
RT	"Cloning and sequencing of the genes of 2-hydroxyglutaryl-CoA	
RT	dehydratase from <i>Acidimicrococcus fermentans</i> .";	
RL	Eur. J. Biochem. 181:741-746(1989).	
RN	[3]	
RP	CHARACTERIZATION.	
RC	STRAIN=ATCC 25085;	
RX	MEDLINE=95331308; PubMed=7607244;	
RA	Mueller U., Buckel W.;	
RT	"Activation of (R)-2-hydroxyglutaryl-CoA dehydratase from	
RT	<i>Acidimicrococcus fermentans</i> .";	
RL	Eur. J. Biochem. 230:698-704(1995).	

NT crystal structure of the *Acetamulococcus* ferredoxin 2;  
RL hydroxyglutaryl-CoA dehydratase component A. " ;  
J. Mol. Biol. 307:297-308(2001).  
CC 1- FUNCTION: REQUIRED FOR THE ACTIVATION OF (R)-2-HYDROXYGLUTARYL-CoA  
CC DEHYDRATASE. THIS PROTEIN IS EXTREMELY SENSITIVE TOWARDS OXYGEN.

CC -1- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: TO E. COLI YJIL AND M. JANNASCHII M70004 AND MJO800.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: X59645; CAA42196.1; -  
 CC EMBL: X14252; CAA32464.1; -  
 CC PIR: S04476.  
 CC PDB: 1HDX; 21-MAR-01.  
 CC InterPro: IPR002731; ATPase\_BadF.  
 CC Pfam: PF01869; BcrAD\_BadFG; 1.  
 CC TIGRFAMs: TIGR00241; CoA\_E\_activ; 1.  
 CC ProDom: PD006344; ATPase\_BadF; 1.  
 CC Iron-sulfur; 4Fe-4S; 3D-structure.  
 CC METAL 127 127 IRON-SULFUR (4FE-4S).  
 CC METAL 166 166 IRON-SULFUR (4FE-4S).  
 CC CONFLICT 196 196 A -> P (IN REF. 2).  
 CC CONFLICT 209 209 V -> L (IN REF. 2).  
 CC CONFLICT 214 224 MTGVAQNYGV -> HDRCSPPELWL (IN REF. 2).  
 CC SEQUENCE 260 AA; 27269 MW; 7E97044DBE805AC CRC64;  
 CC  
 CC Query Match 6.9%; Score 90; DB 1; Length 260;  
 CC Best Local Similarity 25.4%; Pred. No. 3.8;  
 CC Matches 48; Conservative 29; Mismatches 58; Indels 54; Gaps 11;  
 CC  
 CC QY 89 GENAKIGI-DVREKPSAGADRLVNAIGAAMYVPLVVID-----SGTATT 135  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 68 GRNSLEGIAADKMSLS-----CHMGASFTWPNVHTVIDIGGDQVKVHVENGPMTN 120  
 CC QY 136 FDI-----VAADGAFEGGIITAPGINSQALHEAAK-LPRIAIOIP-----AGNRIV----- 182  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 121 FQNDKCACTGRFL-DVMAIIEVKVSDLAELGAKSTRKVAISSTCTVFAESEVISOLS 179  
 CC QY 183 -GDTVTSAMOSQFWGYSISIEGLVARIKAEKPEMTVIATGGVASLFEKATSIDHFD 241  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 180 KGTDKIDII-AGTHRSVASRVIGLANRV-----GIVKDVMTGGVA-----Q 220  
 CC QY 242 DLTIRGLE 250  
 CC DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 221 NYGVGGALE 229  
 CC  
 CC RESULT 9  
 CC -1- PE\_RHISN  
 CC ID NIFE\_RHISN STANDARD; PRT; 496 AA.  
 CC AC P55673;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Nitrogenase iron-molybdenum cofactor biosynthesis protein nife.  
 CC GN NIFE OR Y4VN.  
 CC OS Rhizobium sp. (strain NGR234).  
 CC OG Plasmid sym pNGR234a.  
 CC OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 CC OC Rhizobiaceae; Rhizobium.  
 CC OX NCBI\_TaxID=394;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE-97305956; PubMed-9163424;  
 CC RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 CC RA Perret X.;  
 CC RT Molecular basis of symbiosis between Rhizobium and legumes.\*;  
 CC RL Nature 387:394-401(1997).  
 CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF  
 CC THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).  
 CC -1- PATHWAY: Fe-Mo cofactor biosynthesis.

CC -1- SIMILARITY: BELONGS TO THE NIFD/NIFE/NIFK/NIFE/NIFN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: AE000102; AAB91902.1; -  
 CC HSSP: P07328; 3MIN.  
 CC InterPro: IPR000318; Nitrogenase\_compl.  
 CC InterPro: IPR000510; Oxidized\_nitro; 1.  
 CC Pfam: PF00148; Oxidized\_nitro; 1.  
 CC TIGRFAMs: TIGR01283; nife; 1.  
 CC PROSITE: PS00699; NITROGENASE\_1\_1; 1.  
 CC PROSITE: PS00090; NITROGENASE\_1\_2; 1.  
 CC Nitrogen fixation; Plasmid.  
 CC KW SEQUENCE 496 AA; 54793 MW; D78472D8F5410A3A CRC64;  
 CC  
 CC Query Match 6.9%; Score 90; DB 1; Length 496;  
 CC Best Local Similarity 25.4%; Pred. No. 8;  
 CC Matches 46; Conservative 30; Mismatches 67; Indels 38; Gaps 8;  
 CC  
 CC QY 61 IISVVPO-----SIFNLRLSRRYFNVPLVIGENAKLGDVRIKPS-----AG 107  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 183 VIGTVEPDGAGPCDINTILGEFNLWPLKLL-----ERLGIRVACIPGDARYLDIAS 238  
 CC QY 108 ADRLVNAIGAAMYVPLVVIDSGTATTFDIVAADGAFEGGIITAPGINSQALHEAAK 167  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 239 AHR-----ARAAMLVCGSTALINLARKMEERWDIPFFEGSYG-----ITATSEALRQ 285  
 CC QY 168 LPRIAIORPAGNRIVG-TDVTSAMOSQFWGYSISIEGLVARIKAEKPEMTVIATGGVA 226  
 CC DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
 CC 286 IADLVKKGTDLEILDRDALLAEEAIAW---KKLEERYRPLKGR-----VLNTGGVK 338  
 CC QY 227 S 227  
 CC DB |  
 CC 339 S 339  
 CC  
 CC RESULT 10  
 CC ID DNAK\_YERPE STANDARD; PRT; 636 AA.  
 CC AC Q8ZIM7;  
 CC DT 15-JUN-2002 (Rel. 41, Created)  
 CC DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Chapterone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
 CC protein) (HSP70).  
 CC GN DNAK OR YPO0468.  
 CC OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC OC Yersinia pestis.  
 CC OC Yersinia.  
 CC OX NCBI\_TaxID=632;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN-CO-92 / Biovar Orientalis;  
 CC RX MEDLINE-21470413; PubMed-11586360;  
 CC RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 CC RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 CC RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,  
 CC RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 CC RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 CC RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 CC RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
 CC RT Genome sequence of Yersinia pestis, the causative agent of plague.\*;  
 CC RL Nature 413:523-527(2001).  
 CC -1- FUNCTION: Acts as a chaperone (By similarity).  
 CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.



CC PANCREAS: EXPRESSED AT LOWER LEVELS IN PLACENTA AND HEART. VERY

NA Gadel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hildebrandt A.,

RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT \*The genome sequence of Schizosaccharomyces pombe\*;  
RL Nature 415:871-880(2002).  
RN [2]  
RP SEQUENCE OF 269-556 FROM N.A.  
RC STRAIN=PR745;  
RX MEDLINE=98162722; PubMed=9501991;  
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;  
RT \*Identification of open reading frames in Schizosaccharomyces pombe  
cDNAs\*;  
RL DNA Res. 4:363-369(1997).  
CC -|- SIMILARITY: BELONGS TO THE CGI-48 FAMILY OF WD-REPEAT PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AL022299; CAA18383.1;  
DR EMBL; D89098; BAA13761.1;  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 4.  
DR SMART; SM00320; WD40; 4.  
DR PROSITE; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
DR PROSITE; PS50082; WD\_REPEATS\_2; 1.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
FT REPEAT 216 254  
FT REPEAT 259 298 WD 1.  
FT REPEAT 299 329 WD 2.  
FT REPEAT 330 429 WD 3.  
FT REPEAT 438 479 WD 4.  
FT REPEAT 485 523 WD 5.  
SQ SEQUENCE 556 AA; 62715 MW; 2DF02D921F9D5D7D CRC64;  
  
Query Match 6.9%; Score 89; DB 1; Length 556;  
Best Local Similarity 21.5%; Pred. No. 11;  
Matches 52; Conservative 35; Mismatches 101; Indels 54; Gaps 9;  
  
23 VAQWRSATSTTADEYVWVLSQSLSMOGLGFRAIDAVIISVVPQSFNRLNRRYPEN 82  
150 VPEWAKKQDVTEDEFE-----NALSEKSVIPKSLKSLFKSSVSYIN 191  
83 VEPLVIGE---NAKLGIDVRIKPSKPSAGADRLVNAIGAMVTPGLVIVDSTATTEDIV 139  
192 QSSKLLAFPTGINKRLKANFOAPSHSG-----IRCMSIHYPFPLLLTCGFDRLTRY 244  
140 RADGAFEGGIIAPGNLSMOALHEAAKLPRTAIQRPAGNRIVGDTVSAMQSGVFWGYI 199  
245 QLDGK-----VNPVTSUHLKSSAL-QTALFHPDGKRVIAAGRRKYM---YIWDL 291  
200 SLIEGLVARIAERG-----EPMVTIATGGVASLFEGATDSID-----HPDSLITIG 247  
292 SAQVKVSRMYQENQFQPSMERFHYDPTGKYIAL-EGRSGHNLHLALTGQATSEFKIEG 350  
248 LL 249  
351 VL 352  
  
RESULT 13  
LEU2\_RHILLO

ID LEU2\_RHILLO STANDARD; PRT; 469 AA.  
AC Q98EF1;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE 3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)  
DE (3-isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).  
GN LEUC OR MLL4272.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT \*Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti\*;  
RL DNA Res. 7:331-338(2000).  
CC -|- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate  
CC and 3-isopropylmalate, via the formation of 2-isopropylmaleate.  
CC -|- CATALYTIC ACTIVITY: 3-isopropylmaleate = 2-isopropylmaleate +  
CC H(2)O.  
CC -|- CATALYTIC ACTIVITY: 2-isopropylmaleate + H(2)O = 2-  
CC isopropylmaleate.  
CC -|- PATHWAY: Leucine biosynthesis; second step.  
CC -|- SUBUNIT: Heterodimer of leuc and leucD (By similarity).  
CC -|- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY. LEUC 1  
CC SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AP003003; BAB50968.1;  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF004430; LeucC.  
DR Pfam; PF00330; aconitase; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMS; TIGR00170; leuc; 1.  
DR PROSITE; PS00450; ACONITASE\_1; 1.  
DR PROSITE; PS01244; ACONITASE\_2; 1.  
KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 350 350  
FT METAL 410 410 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 469 AA; 50871 MW; 641532D3D4F06888 CRC64;  
  
Query Match 6.8%; Score 88.5; DB 1; Length 469;  
Best Local Similarity 27.8%; Pred. No. 9.8;  
Matches 54; Conservative 26; Mismatches 69; Indels 45; Gaps 13;  
  
64 SYVPSIFNLRLNRRYFNVEPLVIGENAK-LGIDVRIKPSKPSAGADRLVNAIG--AAMV 120  
68 STSPERKFGIKNEER---IQVELAKNAKDFGEVYSENDIRQG---IVHIGPEQGT 121  
121 YPGPLVVI-DSGTATTFDVAADGAFEGGIIAPGNLSMOALHEAAKLPRTAIQRPAGN 179  
122 LPGMTIVCGDSHTST-----HGAF--GALAHGIGTS-EVEHVLAT---QTLIQKAKN 168  
180 RIVGDTVTSAMQSGVFWGYISLIEGLVAR--IKAERGEPMVTIATGGVASLFEGATDSID 237  
169 MLVRVDG-----QLPEGVTAKDIIAIIIGE---IGTAGGTGVYVYAGEAI- 211

```
QY 238 HFDSDLIRGLLEI 251
Db 212 ---SSMEGRMTI 222

RESULT 14
PGK_RHIME
ID PGK_RHIME STANDARD; PRT; 399 AA.
AC Q92M79;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Phosphoglycerate kinase (EC 2.7.2.3).
GN PGK OR R02765 OR SMC03981.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
NCBI_TaxID=382;
X 1
N 1
AP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Bohn D., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsberger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021".
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
CC phospho-D-glyceroyl phosphate.
CC -1- PATHWAY: Second phase of glycolysis; second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL591791; CAC47344.1;
CC InterPro; IPR001576; PGK.
CC Pfam; PF00162; PGK; 1.
CC PROSITE; PS00111; POLYMERASE_KINASE; 1.
CC TRANSFERASE; Kinase; Glycolysis; Complete proteome.
CC KW SEQUENCE 399 AA; 41909 MW; 96838633CF000126 CRC64;
CC
CC Query Match 6.8%; Score 88; DB 1; Length 399;
CC Best Local Similarity 25.0%; Pred. No. 8.9;
CC Matches 58; Conservative 25; Mismatches 65; Indels 84; Gaps 9;
QY 50 QGLG--FRAIDAVI--SSVVPQISFNRLNLSRRYFNVEPLVIG----- 89
Db 181 KGLGNRPVVAIVGKAVSKIDLLQNLVK---VDALVIGGGMANTFLAAQGVGVCKS 237
QY 90 -----ENAK-----LGIDVRIKPSAGADRLVNAIGAAVMYVPGPLVVI 128
Db 238 LCEHDLAETAKAILAAASEAGCAIVLPVDGVVAREFKAGADNEVVDIKA---IPADAMVL 294
QY 129 DSG-----TATTFDIVADGAFEGITAPGINLSMQALHEAAKLPRITAIORP 176
Db 295 DVGPKSIEAINEWISRAETFLWNGPLGAFE---IAPFDRATVAAAKHAAARTRAGSLVSV 351
QY 177 AGNRIVGDTTYSAMQSGVFWFGYISLIEGLVARIKAERCEPMTVTATGCVASL 228
Db 352 AG-----GGDTVAALN-----HAEVADDTTYVSTAGGAF 381

RESULT 15
BCC4_ACEXY
ID BCC4_ACEXY STANDARD; PRT; 1307 AA.
AC Q9WX71;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cellulose synthase 2 operon protein C precursor.
GN BCS11.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_TaxID=28448;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 7664 / IFO 13693;
RX PubMed=10382968;
RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
RA Inoue Y.;
RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
RT 7664: implication of a novel set of cellulose synthase genes.";
RL DNA Res. 6:109-115(1999).
CC -1- FUNCTION: Required for maximal bacterial cellulose synthesis (By
CC similarity).
CC -1- PATHWAY: Bacterial cellulose biosynthesis.
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: BELONGS TO THE ACSC/BCSC FAMILY.
CC -1- SIMILARITY: CONTAINS 9 TPR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB015803; BAA7596.1;
CC InterPro; IPR001440; TPR.
CC Pfam; PF00515; TPR; 5.
CC SMART; SM00028; TPR; 3.
CC Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
CC Signal.
CC SIGNAL 1 55
CC CHAIN 56 1307
CC REPEAT 97 130
CC REPEAT 270 303
CC REPEAT 339 372
CC REPEAT 374 406
CC REPEAT 458 491
CC REPEAT 493 525
CC REPEAT 528 561
CC REPEAT 754 787
CC REPEAT 788 821
CC SEQUENCE 1307 AA; 141286 MW; 6F5A5F38AD9FE53 CRC64;
CC
CC Query Match 6.8%; Score 88; DB 1; Length 1307;
CC Best Local Similarity 23.2%; Pred. No. 35;
CC Matches 55; Conservative 40; Mismatches 96; Indels 46; Gaps 10;
QY 53 GFRAIDAVI---SSVVPQISFNRLNLSRRYFNVEPLVIGENAKLIGIDVRIKPSAGAD 109
Db 934 QAQAMDROIYLRDSVSPQ--FDANTFVRSRTGAGL--GQLTEFAVPITATLPFESWDH 989
QY 110 RLVNAIGAAVMYVPG-PLVVIDSGTATTFDIVADGAFEGG---LIAPGINLSMQALH--- 162
Db 990 RLFSVYTTLLFTGDPILT--NAVSAHQFGFTVAVNGARPMGCHYHYTQGVGLSLNYVNRWF 1047
QY 163 -----EAAAKLPRIATIORPAGNRIVGTDTVSAMQ-----SGVFW 196
Db 1048 AADVGSSPLGFPITNVVGGLEFAPRLTRNLGLRISGRRMVTDSLSYAGERDPGCKLW 1107
```

QY 197 GYISLIEGLVARIKAERGEPMTVIATGGVASLFEAGATDSIDHEDSDLTIRGLLEIYR 253  
| : : | | | | | | | | | : : : | : : :  
Db 1108 GGVTRLFHGHCALEWSARG--WNAYAGGFAYL--GGTNVIGNTEAGAGGSATVWQ 1160

Search completed: June 24, 2003, 22:11:50  
Job time : 6.23014 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

. OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 22.8708 Seconds  
(without alignments)  
2342.388 Million cell updates/sec

Title: US-09-813-453A-51  
Perfect score: 1299  
Sequence: 1 MLLAIEQGNNTMFAIHGCA.....SOLTINGLLEIYRNTIAES 260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

--Searched: 671580 seqs, 206047115 residues 671580  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp-archaea.\*
- 2: sp-bacteria.\*
- 3: sp-fungi.\*
- 4: sp-human.\*
- 5: sp-invertebrate.\*
- 6: sp-mammal.\*
- 7: sp-mhc.\*
- 8: sp-organelle.\*
- 9: sp-phage.\*
- 10: sp-plant.\*
- 11: sp-rodent.\*
- 12: sp-virus.\*
- 13: sp-vertebrate.\*
- 14: sp-unclassified.\*
- 15: sp-rviro.\*
- 16: sp-bacteriap.\*
- 17: sp-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB ID	Description
		Match	Length			
1	1299	100.0	261	16	Q9A6Z1	Q9A6Z1 caulobacter
2	598	46.0	251	16	Q9KGH5	Q9KGH5 bacillus ha
3	568	43.7	258	2	Q9F985	Q9F985 bacillus st
4	551	42.4	259	16	Q8XHL5	Q8XHL5 clostridium
5	530	40.8	255	16	Q8R7M2	Q8R7M2 thermococ
6	523	40.3	259	16	Q9YAC5	Q9YAC5 listeria mo
7	519	40.0	273	16	Q97EB4	Q97EB4 clostridium
8	517	39.8	265	16	Q9X8N6	Q9X8N6 streptomyce
9	516	39.7	259	16	Q9ZFE4	Q9ZFE4 listeria in
10	423	32.6	256	16	Q8RFE4	Q8RFE4 fusobacteri
11	396.5	30.5	262	16	Q9RX54	Q9RX54 deinococcus
12	389.5	30.0	274	16	Q9CD56	Q9CD56 mycobacteri
13	382.5	29.4	272	16	Q96282	Q96282 mycobacteri
14	359.5	27.7	246	16	Q9WZY5	Q9WZY5 thermotoga
15	337.5	26.0	273	16	Q83446	Q83446 treponema p
16	307.5	23.7	212	2	Q32514	Q32514 desulfovibr

17	201	15.5	295	16 Q8Y2M4	Q8Y2M4 ralstonia s
18	189.5	14.6	262	16 Q51477	Q51477 borrelia bu
19	175	13.5	592	16 Q9JW17	Q9JW17 neisseria m
20	172	13.2	592	16 Q9JXF1	Q9JXF1 neisseria m
21	163.5	12.6	276	16 Q8YQD7	Q8YQD7 anabaena sp
22	160	12.3	248	16 Q9HWC1	Q9HWC1 pseudomonas
23	147	11.3	257	16 P74045	P74045 synecocyst
24	140.5	10.8	242	16 Q9PC14	Q9PC14 xylella fas
25	125.5	9.7	229	16 Q67753	Q67753 aquifex aeo
26	111	8.5	209	16 Q9PIA9	Q9PIA9 campylobact
27	107.5	8.3	401	16 Q910F1	Q910F1 pseudomonas
28	102.5	7.9	1889	16 Q9ZJV3	Q9ZJV3 rhizobium m
29	99.5	7.7	729	2 P96168	P96168 vibrio harv
30	99.5	7.7	980	10 Q949G8	Q949G8 malus flori
31	98	7.6	392	16 P96214	P96214 mycobacteri
32	98	7.5	1536	2 Q48031	Q48031 haemophilus
33	97.5	7.5	456	2 Q54101	Q54101 saccharopol
34	97.5	7.5	5291	16 Q8X2T1	Q8X2T1 escherichia
35	96.5	7.4	5188	16 Q8X4H5	Q8X4H5 escherichia
36	96	7.4	460	16 Q8YQT7	Q8YQT7 anabaena sp
37	95.5	7.4	1787	16 Q8G329	Q8G329 mycobacteri
38	95	7.3	388	16 Q9XZS1	Q9XZS1 streptococc
39	94.5	7.3	4735	2 Q54666	Q54666 amycolatops
40	94.5	7.3	9507	2 Q9EWAI	Q9EWAI streptomyce
41	94	7.2	530	16 Q8UK38	Q8UK38 agrobacteri
42	92.5	7.1	372	2 Q8RIU8	Q8RIU8 corynebacte
43	92.5	7.1	496	10 Q9S833	Q9S833 arabidopsis
44	92	7.1	384	16 P74281	P74281 synecocyst
45	92	7.1	416	16 Q93J85	Q93J85 streptomyce

ALIGNMENTS

RESULT 1

Q9A6Z1 ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.  
AC Q9A6Z1;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE Transcriptional activator, putative, Baf family.  
GN CC1935.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,  
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Barry K.,  
Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
DR EMBL; AE005867; AAK23910.1; -;  
DR TIGR; CC1935;  
DR InterPro; IPR004619; Baf.  
DR Pfam; PF03309; Bvg\_acc\_factor; 1.  
DR TIGRfam; TIGR00671; baf; 1.  
KW Complete proteome.  
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 100.0%; Score 1299; DB 16; Length 261;  
Best Local Similarity 100.0%; Pred. No. 8.4e-89;  
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLAIEQGNNTMFAIHGASWAQWRSAESTRTADEYVWVLSQLSMOGLGFRAIDAV 60

[illegible]

```

DR InterPro; IPR004619; Baf.
DR InterPro; IPR00515; BPD_transp.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00402; BPD_TRANS_P_NN_MEMBER; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;

Query Match 42.4%; Score 551; DB 16; Length 259;
Best Local Similarity 42.7%; Pred. No. 2.8e-33;
Matches 111; Conservative 53; Mismatches 88; Indels 8; Gaps 3;

QY 1 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60
DB 1 MILLIDVGNITVIGIHNDNEKXIASWRISTSKTSDEYSIQVMQLFNQAKLPEDVDEGI 60

QY 61 IISSVVPQISFNLRNLSRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAAVY 120
DB 1 IISSVVPNIMHSLNENVRKCKEPIVVGPGIKTGINIKYDNPKEVGADRVNVAFAFEK 120

QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIOPAGNR 180
DB 1 HKPMIILIDFTATTFCAITEKGDYLGNGICPGIISADALFERRAKLPRIELEKPP--KS 178

QY 181 IVGDTVTSAMOSGVFWGYSISIEGLVARIKAE-----RGEPTVIATGGVSLFEGATDS 235
DB 179 VICNTVTSMAQGIYIGVIGVEYIVRKKEMMDLGEKEPP-VLATGGLAKLVYSETDV 237

QY 236 IDHFDSDLTIRGLEIYRRN 255
DB 238 IDEVDRKLTLEGLKLYEKN 257

RESULT 5
Q8R7M2 PRELIMINARY; PRT; 255 AA.
AC Q8R7M2
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DE Putative transcriptional regulator, homologs of Bvg accessory
DE factor.
GN TTE2381.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
P [1]
P SEQUENCE FROM N.A.
AC STRAIN-MBAT / JCM11007;
RX MEDLINE-21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE031380; AAM25520.1;
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 255 AA; 27816 MW; C3C620ECBC8CA6ED CRC64;

Query Match 40.8%; Score 530; DB 16; Length 255;
Best Local Similarity 40.8%; Pred. No. 1e-31;
Matches 104; Conservative 59; Mismatches 90; Indels 2; Gaps 1;

QY 1 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60
DB 1 MLLAFDVGNTIVMGVFGKLLHSFRISTDKNTYDEYGLVNLQILYNGISLFEIDDV 60

QY 61 IISSVVPQISFNLRNLSRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAAVY 120
DB 1 IISSVVPPLMTLQVMSLKYPTKPIVVGPGIKTGINIKYDNPKEVGADRVNVAFAVEL 120

QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIOPAGNR 180

DB 179 VICNTVTSMAQGIYIGVIGVEYIVRKKEMMDLGEKEPP-VLATGGLAKLVYSETDV 237
DB 238 IDEVDRKLTLEGLKLYEKN 257

DB 241 SDLTIRGLEIYRRN 255
DB 239 EMLTLEGLRIIYERN 253

DB 121 YGGPVIIVIDFGTATTFCAISEKGEYLGIIAAGLMSADALFQRTAKLPKIDLTAKPP--T 178
DB 181 IVGDTVTSAMOSGVFWGYSISIEGLVARIKAEGBEPTMTVIATGGVSLFEGATDSIDHFD 240
DB 179 VINRTVASMOSGIYGHVGVYIVTRMKGEFAPFSAVVVATGGFANMAIESKTIIDTVN 238

QY 241 SDLTIRGLEIYRRN 255
DB 239 EMLTLEGLRIIYERN 253

RESULT 6
Q8YAC5 PRELIMINARY; PRT; 259 AA.
AC Q8YAC5
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DE Hypothetical protein lmo0221.
GN LMO0221.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EGD-E / SEROVAR 1/2A;
RX MEDLINE-21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusnlok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant A., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-Lopez N., Hain T., Hauf J., Jackson D.,
RA Gautier L., Goebel W., Gomez-Lopez N., Kuhn M., Kunst F., Kurapkai G.,
RA Jones L.-M., Kaerst U., Kreft J., Madueno A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591974; CAD00748.1;
DR MEROPS; M41.009;
DR ListList; LMO00221;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 40.3%; Score 523; DB 16; Length 259;
Best Local Similarity 39.2%; Pred. No. 3.4e-31;
Matches 100; Conservative 59; Mismatches 94; Indels 2; Gaps 1;

QY 1 MLLAIEQNTNTMFAIHGASWVAQWRSATSTRTADEYVYVWLSQLLSMOGLGFRAIDAV 60
DB 1 MILVIDVGNITVIGVYEKQKLLKHWRTTDRHTSDGLMGLVLFNFYSANLTPSDIQGI 60

QY 61 IISSVVPQISFNLRNLSRRYFNVEPLVIGENAKLIGIDVRIEKPSEAGADRLVNAIGAAVY 120
DB 61 IISSVVPPIIMHMETVCVRYFNIRPLIVGPGIKTGLNLKVDNPREGSRIVNVAASEE 120

QY 121 YPGPLVVIDSGTATTFDVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAIOPAGNR 180
DB 121 YGTPVIVVDFTATTFYIDESGVYQGAIAFGIMISTEALYNRAAKLPVDIAE--SSQ 178

QY 181 IVGDTVTSAMOSGVFWGYSISIEGLVARIKAEGBEPTMTVIATGGVSLFEGATDSIDHFD 240
DB 179 IIGKSTVSSMAQGIYGFVQCEGIIAEMKKQSNAPVVVATGGTGLARMITKSSAVDILD 238

QY 241 SDLTIRGLEIYRRN 255
DB 239 EMLTLEGLRIIYERN 253

```





```

01-DEC-2001 (TREMBLrel. 19, Last sequence update)
01-JUN-2002 (TREMBLrel. 21, Last annotation update)
Hypothetical protein lin0253.
GN LIN0253.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Franquet L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui A., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durand L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596164; CAC95486.1; -.
DR Listlist; LIN00253; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28227 MW; 554B03A0C0EFA64F CRC64;

Query Match 39.7%; Score 516; DB 16; Length 259;
Best Local Similarity 39.2%; Pred. No. 1.1e-30;
Matches 100; Conservative 58; Mismatches 95; Indels 2; Gaps 1;

QY 1 MLLAEOGNTMTFAIHGASWVAQWSATSTRTADEYVWVLSQLSMOGLGFRAIDAV 60
DB 1 MILVIDGNTCTGVYKEQLLRHWRMTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60
QY 61 IISVWPQSIPLNLSRRYFNVPLVIGENAKL-----GIDVRIKPSAGADRLVN 120
DB 61 IISVWPPIHMTMTCVRYENIRPLVIGPKTGLNKLKVDNPREIGSDRLVNAVAASEE 120
QY 121 YPGVLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRIAPAGNR 180
DB 121 YGTPVIVDFCTATTCFYIDEAGVYQGGAIAPGIMISTEALYNRAKLPVRYDIAE--SSQ 178
QY 181 IVGTDTSVMSQSGVFWGYSILIEGLVARIKAEKPEMTVIATGGVASFEGATSDIDHFD 240
DB 179 IIGKSTVMSQAGIFYGFIGOCEGIIAEMKKQSNTPSPVVATGGLARMITEKSSAVDILD 238
QY 241 SLDLIRGLLEIYRN 255
DB 239 PFLTKGLLEIYRN 253

RESULT 10
Q8RFE4 PRELIMINARY; PRT; 256 AA.
AC Q8RFE4;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bvg accessory factor.
GN FN0761.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;

RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fostein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010586; AAL94957.1; -.
DR KW Complete proteome.
SQ SEQUENCE 256 AA; 28601 MW; CDF681127F08914B CRC64;

Query Match 32.6%; Score 423; DB 16; Length 256;
Best Local Similarity 34.2%; Pred. No. 8.9e-24;
Matches 90; Conservative 61; Mismatches 96; Indels 16; Gaps 6;

QY 1 MLLATEOGNTMTFAIHD-GASWVAQWSATSTRTADEYVWVLSQLSMOGLGFRAIDA 59
DB 1 MIIGIDIGNTHIVTGIYDNGNELISTFRIATNDKMTDEYFYSFNITKYNEISIKKVA 60
QY 60 VIISVWPQSIPLNLSRRYFNVPLVIGENAKL-----GIDVRIKPSAGADRLVN 113
DB 61 ILISSVVPNIITFFQFARKYKVEATIVDLKKLPFTFAKGINY-----TGFADRIID 115
QY 114 AIGAAVYVPGP-LVVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMOALHEAAKLPRIA 172
DB 116 ITEAMQKYPDKNLVIFDFGTATTYDVL-KKGVIYGGGILPGIDMSINALYGNATKLPVK 174
QY 173 IORPAGNRIVGTDTSVMSQSGVFWGYSILIEGLVARIKAEKPEMTVIATGGVASFEGA 232
DB 175 FTFPSS--VLGTDTKQIAAIFFGYAGQIKHRIKINEELNEIFEIVLATGGLGKILSAE 232
QY 233 TDSIDHFDSDLTIRGLLEIYRN 255
DB 233 IDEIDYDANLSLKGILTYLKLN 255

RESULT 11
Q9RX54 PRELIMINARY; PRT; 262 AA.
AC Q9RX54;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein DR0461.
GN DR0461.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001905; AAF10040.1; -.
DR TIGR; DR0461; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001220; Lectin_legB.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR KW Hypothetical protein; Complete proteome.

```

SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78795A0 CRC64;

Query Match 30.5%; Score 396.5; DB 16; Length 262;  
 Best Local Similarity 36.1%; Pred. No. 8.5e-22;  
 Matches 92; Conservative 45; Mismatches 109; Indels 9; Gaps 5;

QY 2 LLAIQGNNTMFAIHGASWVAQ--WRSATSESTRADVEYVWLSQGLGFRDAIDAV 60  
 DB 6 LLAVDIGNTTVLGLADAGSLTHTWRITNREMLPDDLALQLHLFTLAGAPIR--AA 63  
 QY 61 IISVVPOSIFNRLNLSRYFNVEPLVIGENAKLGDIVRIKPSBAGADRLVNAIGAAMV 120  
 DB 64 VLSSVAPPVGENYALAKRHFWIDAFVSAENLPDVTVELDTPGSGVADRLCNLEGAEK- 122  
 QY 121 YGPIPL---VWIDSGPATTFFDIVADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIORPA 177  
 DB 123 YLGLDYAVVVDFTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRIITLQAP- 181  
 QY 178 GNRIVGDTVSAMQSGVFWYISLIEGLVARIKARGEPMVTIATGGVASLPEGATDSID 237  
 DB 182 -ETAIGKNTVHALQSLVFGYAEWVDGLRLRRIAELPGEAVAVATGFSRTVQGICQSID 240  
 QY 238 HFDSDLTIRGLLEIY 252  
 DB 241 YYDETLLRLGLVELW 255

RESULT 12

ID 09CD56 PRELIMINARY; PRT; 274 AA.

AC 09CD56;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein ML0232.

GN ML0232.

OS Mycobacterium leprae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1769;

RP [1]

RC STRAIN-TN;

RC STRAIN=21128732; PubMed=11234002;

RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus."

RL Nature 408:1007-1011(2001).

DR EMBL; AL583917; CAC29740.1; -.

DR Leproma; ML0232; -.

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg\_acc.factor; 1.

DR TIGRFAMs; TIGR00671; baf; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 30.0%; Score 389.5; DB 16; Length 274;  
 Best Local Similarity 34.3%; Pred. No. 3e-21;  
 Matches 93; Conservative 53; Mismatches 108; Indels 17; Gaps 5;

QY 1 MLLAIEQGNNTMFAIHG-----ASWVAQWRSATSESTRADVEYVWLSQGLGFR 56  
 DB 1 MLLAIDVRNTHTVVGLSGSKEHAKVVOQWRIRTSEVTADALALIDGLI---GDSER 57  
 QY 57 IDAVIISVVPQSIFNRLNLSRYFNVEPLVIGENAKLGDIVRIKPSBAGADRLVNAI 115

SQ SEQUENCE 262 AA; 27839 MW; 965EAD2F78795A0 CRC64;

Query Match 30.5%; Score 396.5; DB 16; Length 262;  
 Best Local Similarity 36.1%; Pred. No. 8.5e-22;  
 Matches 92; Conservative 45; Mismatches 109; Indels 9; Gaps 5;

QY 2 LLAIQGNNTMFAIHGASWVAQ--WRSATSESTRADVEYVWLSQGLGFRDAIDAV 60  
 DB 6 LLAVDIGNTTVLGLADAGSLTHTWRITNREMLPDDLALQLHLFTLAGAPIR--AA 63  
 QY 61 IISVVPOSIFNRLNLSRYFNVEPLVIGENAKLGDIVRIKPSBAGADRLVNAIGAAMV 120  
 DB 64 VLSSVAPPVGENYALAKRHFWIDAFVSAENLPDVTVELDTPGSGVADRLCNLEGAEK- 122  
 QY 121 YGPIPL---VWIDSGPATTFFDIVADGAFEGGIIAPGINLSMOALHEAAAKLPRIAIORPA 177  
 DB 123 YLGLDYAVVVDFTSTNFDVVGRRFLGGILATGAQVSADALFARAAKLPRIITLQAP- 181  
 QY 178 GNRIVGDTVSAMQSGVFWYISLIEGLVARIKARGEPMVTIATGGVASLPEGATDSID 237  
 DB 182 -ETAIGKNTVHALQSLVFGYAEWVDGLRLRRIAELPGEAVAVATGFSRTVQGICQSID 240  
 QY 238 HFDSDLTIRGLLEIY 252  
 DB 241 YYDETLLRLGLVELW 255

RESULT 13

ID 006282 PRELIMINARY; PRT; 272 AA.

AC 006282;

DT 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical 29.3 kDa protein (Transcriptional activator, putative,  
 Baf family).

DE Baf family).

GN RV3600C OR MTCY07H7B.22 OR MT3706.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RP [1]

RC STRAIN=H37RV;

RC MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton J., Squares S., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."

RL Nature 393:537-544(1998).

RN [2]

RC STRAIN=CDC 1551 / OSHKOSH;

RC STRAIN=CDL 1551 / OSHKOSH;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z95557; CAB08944.1; -.

DR EMBL; AE007170; AAK48063.1; -.

DR TIGR; MT3706; -.

DR Tuberculist; RV3600C; -.

DR InterPro; IPR004619; Baf.

DR Pfam; PF03309; Bvg\_acc.factor; 1.

DR TIGRFAMs; TIGR00671; baf; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 272 AA; 29304 MW; 5D70E6E0F09AC8B CRC64;

Query Match 29.4%; Score 382.5; DB 16; Length 272;  
 Best Local Similarity 34.0%; Pred. No. 9.8e-21;  
 Matches 90; Conservative 53; Mismatches 107; Indels 15; Gaps 5;

QY 1 MLLAIEQGNNTMFAIHG-----ASWVAQWRSATSESTRADVEYVWLSQGLGFR 56  
 DB 1 MLLAIDVRNTHTVVGLSGSKEHAKVVOQWRIRTSEVTADALALIDGLI---GDSER 57

```

QY 57 IDAVLISSVPSQIFNLRLNRRYF-NVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAI 115
DB 58 LTGTAALSTVPSVLEHVRIMLDQYWPSPVPHVLIIEPGVRTGIPLLVDNPKVGADRIVNCL 117
QY 116 GAAMVYGPPLVVIDSGTATTFDIYAADGAFEGGIAPGINLSMOALHEAAAKLPRIATQR 175
DB 118 AAYDRFKAAIVDFGSSICVDVVSARKGEGFAGGAIAPGVQVSSDAAAARSALRRVELAR 177
QY 176 PAGNRIVGTDTVSAMQSVFNGYISLIEGLVARIKAEGR-----EPMVTIATGVASLIFE 230
DB 178 P--RSVVGNKTVCEKQAGAFEGFAGLVDGLVGRIREDSVGFSDHVAIVATGHTAPLL 235
QY 231 GATDSIDHFDSDLTIRGLLEIYRN 255
DB 236 PELTVDRHYDQHLTLOGLRLVFERN 260

RESULT 14
-Q9WZY5
D Q9WZY5 PRELIMINARY; PRT; 246 AA.
AC Q9WZY5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TW0883.
GN TW0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
DR EMBL; AE001754; AAD35964.1; -
DR TIGR; TW0883; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR -W Hypothetical protein; Complete proteome.
2 SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

Query Match 27.7%; Score 359.5; DB 16; Length 246;
Best Local Similarity 35.5%; Pred. No. 4.4e-19;
Matches 89; Conservative 53; Mismatches 94; Indels 15; Gaps 9;

QY 1 MLLAIEGNTNTMFAIHDGASWVAQWRSATSTRTADEYVWLSQGLGFRAIDAV 60
DB 1 MYLLVDVGNTHSVFSEITDEKGFTRWRSLSTGVTFDELFSLHPL--GDAMREIKGI 57
QY 61 IISWVPSQIFNLRLNRRYFNVPLVIGENAKLG-IDVRIEKPSEAGADRLVNAIGAAM 119
DB 58 GVASVVTQNTVIERFSQKFIHSPIW--KAKNGCVKNVKNFSEVGADRVANVAPVK 115
QY 120 VYGPPLVVIDSGTATTFDIYAADGAFEGGIAPGINLSMOALHEAAAKLPRIATQR 179
DB 116 EYGNGLIIDMTATTVDLV-VNGSYEGGAILPGFMMVHSLFRGTAKPLPLEV-KPA-D 172
QY 180 RIVGTDTVSAMQSVFNGYISLIEGLVARIKAEGRPMTVIATGVASLIFE 238
DB 173 FVVGKDFEENIRLGVNNGSVYALEGIIGRIKEVYGD-LPVVLTGGQSKIVR--DMIKHE 228
QY 239 -FDSDLTIRGL 248
DB 229 IFDEDLIKGV 239

```

```

RESULT 15
O83446
ID O83446 PRELIMINARY; PRT; 273 AA.
AC O83446;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0431.
GN TP0431.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65417.1; -
DR TIGR; TP0431; -
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 273 AA; 28472 MW; 439C9C77CB598BC0 CRC64;

Query Match 26.0%; Score 337.5; DB 16; Length 273;
Best Local Similarity 32.9%; Pred. No. 2.2e-17;
Matches 85; Conservative 53; Mismatches 111; Indels 9; Gaps 5;

QY 2 LLATIEGNTNTMFAI---HDGASWVAQ-WRSATSTRTADEYVWLSQGLGFRAI 57
DB 1 MLLIDVGNHVVFGIGQGGNGRVCRFLRPLADPARKTODEYSLLIHALCERACVGRASL 60
QY 58 DAVIISVYVPSQIFNLRLNRRYFNVPLVIGENAKLGIDVRIEKP--SEAGADRLVNAI 115
DB 61 RDAFISVVVPLTKTIADAVAQISGVQPVVFGFWAYEHLPRVRIPEVPVRAEIGTDLVANAV 120
QY 116 GAAMVYGPPLVVIDSGTATTFDIYAADGAFEGGIAPGINLSMOALHEAAAKLPRIATQR 175
DB 121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVATAPGLRTAVQSLHTGTQPLPLPAL 180
QY 176 PAGNRIVGTDTVSAMQSVFNGYISLIEGLVARIKAEGRPMTVIATGVASLIFE 235
DB 181 P--DSVLKGDTHAVQAGVVRGTLFVIRAMIAQCKELGCRCAAVITGSLRSLFSEVD- 237
QY 236 IDHFDSDLTIRGLLEIYR 253
DB 238 FPPIDAQLTSLGLAHAR 255

Search completed: June 24, 2003, 22:16:14
Job time : 23.8708 secs

```

**THIS PAGE BLANK (USPTO)**